

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 04:30:35 ; Search time 4172.21 Seconds

(without alignments)
11406.588 Million cell updates/sec

Title: US-09-901-484A-179_COPY_3899_4996

Perfect score: 1098
Sequence: 1 gtctctgtctctgtcattt.....agaacttaagttctctgat 1098.

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 3470272 seqs, 21671516995 residues

Word size : 0

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenEmbl:*
1: gb_ba:*
2: gb_hcg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_bcs:*
12: gb_ay:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_ats:*
28: em_un:*
29: em_vl:*
30: em_hcg_hum:*
31: em_hcg_inv:*
32: em_hcg_other:*
33: em_hcg_mus:*
34: em_hcg_pln:*
35: em_hcg_rod:*
36: em_hcg_mam:*
37: em_hcg_vrt:*
38: em_ay:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1098	100.0	56516	6 BD196412	BD196412 Prostatic
2	1098	100.0	56520	6 BD196564	BD196564 Prostatic
3	996	90.7	165799	2 AC100813	AC100813 Homo sapi
4	996	90.7	168136	9 AF287957	AF287957 Homo sapi
5	996	90.7	183485	2 AC022578	AC022578 Homo sapi
6	996	90.7	209718	2 AC091099	AC091099 Homo sapi
7	813	74.0	185834	2 AC009632	AC009632 Homo sapi
8	114	10.4	185834	2 AC009632	AC009632 Homo sapi
9	25	2.3	19087	6 AX345694	AX345694 Sequence
10	24	2.2	6782	6 AX345679	AX345679 Sequence
11	24	2.2	179514	10 AC134447	AC134447 Mus muscu
12	24	2.2	264864	2 AC135128	AC135128 Rattus no
13	23	2.1	47	6 AR290033	AR290033 Sequence
14	23	2.1	47	6 AX114392	AX114392 Sequence
15	23	2.1	6070	6 AX281468	AX281468 Sequence
16	23	2.1	6070	6 AX346581	AX346581 Sequence
17	23	2.1	6070	6 AX348804	AX348804 Sequence
18	23	2.1	6070	6 AX458614	AX458614 Sequence
19	23	2.1	6074	6 AX323760	AX323760 Sequence
20	23	2.1	7238	6 AX345350	AX345350 Sequence
21	23	2.1	15121	6 AX344831	AX344831 Sequence
22	23	2.1	15649	6 AX348975	AX348975 Sequence
23	23	2.1	39403	3 CE0303A3	CE0303A3
24	23	2.1	42740	2 AC090706	AC090706 Homo sapi
25	23	2.1	110394	9 AC003086	AC003086 Homo sapi
26	23	2.1	129109	9 AL390239	AL390239 Human DNA
27	23	2.1	135862	9 AC025757	AC025757 Homo sapi
28	23	2.1	147874	2 AC011278	AC011278 Homo sapi
29	23	2.1	150438	9 AC099517	AC099517 Homo sapi
30	23	2.1	165965	2 AC012287	AC012287 Homo sapi
31	23	2.1	167122	2 AC053532	AC053532 Homo sapi
32	23	2.1	168216	2 AC013398	AC013398 Homo sapi
33	23	2.1	173445	9 AC078876	AC078876 Homo sapi
34	23	2.1	174617	9 AC105180	AC105180 Homo sapi
35	23	2.1	189464	2 AC112194	AC112194 Homo sapi
36	23	2.1	195909	9 AF165124	AF165124 Homo sapi
37	23	2.1	203131	9 AC009070	AC009070 Homo sapi
38	23	2.1	203982	9 AC067854	AC067854 Homo sapi
39	23	2.1	207040	9 AC091984	AC091984 Homo sapi
40	23	2.1	209836	9 AC091926	AC091926 Homo sapi
41	23	2.1	210736	2 AC092792	AC092792 Homo sapi
42	23	2.1	219223	2 BX649211	BX649211 Dantio rer
43	23	2.1	233219	2 AC113253	AC113253 Rattus no
44	23	2.1	252420	2 AE014841	AE014841 Plasmodu
45	23	2.1	253149	2 AC107527	AC107527 Rattus no

ALIGNMENTS

RESULT 1	LOCUS	BD196412	56516 bp	DNA	linear	PAT 17-JUN-2003
BD196412	DEFINITION	Prostatic cancer gene.				
BD196412	ACCESSION	BD196412				
BD196412.1	VERSION	GI:33006182				
UP 2002516657-A/1.	KEYWORDS					
Homo sapiens (human)	SOURCE					
Homo sapiens	ORGANISM					
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	REFERENCE					
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	AUTHORS					
Cohen,D., Blumenfeld,M., Chumakov,I. and Bougueleret,L.	TITLE					
Prostatic cancer gene	JOURNAL					
Patent: JP 2002516657-A 11-JUN-2002;						

Pred. No. is the number of results predicted by chance to have a

[illegible]

FT primer_bind 11622..11639
FT primer_bind 12018..12037
FT primer_bind 11930..11947
FT primer_bind 12339..12358
FT primer_bind 12915..12932
FT primer_bind 13317..13334
FT primer_bind 13216..13233
FT primer_bind 13617..13636
FT primer_bind 13547..13564
FT primer_bind 13962..13981
FT primer_bind 15994..16011
FT primer_bind 16463..16480
FT primer_bind 17304..17324
FT primer_bind 17814..17832
FT primer_bind 18008..18027
FT primer_bind 18423..18442
FT primer_bind 18699..18716
FT primer_bind 19164..19182
FT primer_bind 22589..22609
FT primer_bind 23111..23129
FT primer_bind 25098..25118
FT primer_bind 25657..25674
FT primer_bind 26537..26557
FT primer_bind 27022..27040
FT primer_bind 32262..32281
FT primer_bind 32823..32841
FT primer_bind 34215..34233
FT primer_bind 34624..34644
FT primer_bind 34473..34491
FT primer_bind 34916..34936
FT primer_bind 34702..34722
FT primer_bind 35182..35200
FT primer_bind 35951..35971
FT primer_bind 39971..39991
FT primer_bind 40531..40549
FT primer_bind 40932..40950
FT primer_bind 40629..40649
FT primer_bind 41058..41078
FT primer_bind 45729..45746
FT primer_bind 46186..46203
FT primer_bind 47879..47896
FT primer_bind 48217..48236
FT primer_bind 48902..48922
FT primer_bind 49331..49351
FT primer_bind 49830..49848
FT primer_bind 50271..50290
FT primer_bind 50172..50189
FT primer_bind 50573..50591
FT primer_bind 50541..50560
FT primer_bind 50940..50959
FT primer_bind 50555..50572
FT primer_bind 50964..50983
FT primer_bind 50774..50792
FT primer_bind 51183..51202
FT primer_bind 51146..51165
FT primer_bind 51479..51496
FT primer_bind 51593..51610
FT primer_bind 29734..29744
FT primer_bind 51167..51185
FT primer_bind 51667..51687
FT primer_bind 51387..51403
FT primer_bind 51826..51843
FT primer_bind 51772..51789
FT primer_bind 52199..52217

Query Match 100.0%; Score 1098; DB 6; Length 56520;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTTCTGTCGCGCATTTTATTTATTTGATGAGGAGCAAAATTAATTTCTGT 60
DB 3899 GTTCTGTCGCGCATTTTATTTATTTGATGAGGAGCAAAATTAATTTCTGT 3958

QY 61 TTAGTAACAGAGCGGATTTTGAATTTATAGGGTCTTTTCTGACGTCTGGGTTTC 120
DB 3959 TTAGTAACAGAGCGGATTTTGAATTTATAGGGTCTTTTCTGACGTCTGGGTTTC 4018
QY 121 TGTGTACAAAGGTAAGCTTTCAATATTTTATTTTGTGTTCTGTAAATCAATAG 180
DB 4019 TGTGTACAAAGGTAAGCTTTCAATATTTTATTTTGTGTTCTGTAAATCAATAG 4078
QY 181 AGGAATTAATAGCATCTTCAACATAGACCCAAAGAAAAGATTATAGTATGTC 240
DB 4079 AGGAATTAATAGCATCTTCAACATAGACCCAAAGAAAAGATTATAGTATGTC 4138
QY 241 TGTACCTTATTTTATCTGTGACCTTTGTACATTAACCTTGTCACTGAGTGTGGA 300
DB 4139 TGTACCTTATTTTATCTGTGACCTTTGTACATTAACCTTGTCACTGAGTGTGGA 4198
QY 301 TTAAATTTTATAGCTTCTGTTTCTGTTTGTATAGGACCTCTTTTCTGAAATGTT 360
DB 4199 TTAAATTTTATAGCTTCTGTTTCTGTTTGTATAGGACCTCTTTTCTGAAATGTT 4258
QY 361 TTATACGCTTGTGTTGCAAGGCTAGATGATTCCTGTTCTGTATTAAGTATGTT 420
DB 4259 TTATACGCTTGTGTTGCAAGGCTAGATGATTCCTGTTCTGTATTAAGTATGTT 4318
QY 421 GACTCATTTCTGAAGGAGTTTATGATTTTAAGAGGTTAAGCTTTTAATTAAGT 480
DB 4319 GACTCATTTCTGAAGGAGTTTATGATTTTAAGAGGTTAAGCTTTTAATTAAGT 4378
QY 481 TTATTAATTTATATATTTTAAGAGGATTTTAATTAATTTTATTAATGACATT 540
DB 4379 TTATTAATTTATATATTTTAAGAGGATTTTAATTAATTTTATTAATGACATT 4438
QY 541 TTACACCTTTCACTAGTTTAAAAAATAGTGTTCACAGTAGTCTTGACAGAA 600
DB 4439 TTACACCTTTCACTAGTTTAAAAAATAGTGTTCACAGTAGTCTTGACAGAA 4498
QY 601 TATTTTCTTTTACATAGATTTTATAGCTGAAGGAGTATAGTCCATGATTT 660
DB 4499 TATTTTCTTTTACATAGATTTTATAGCTGAAGGAGTATAGTCCATGATTT 4558
QY 661 ATGATCTGTGCTGGCAGGTAACCTGCTTCAACAAATTAGTGTGATTTCTTGAT 720
DB 4559 ATGATCTGTGCTGGCAGGTAACCTGCTTCAACAAATTAGTGTGATTTCTTGAT 4618
QY 721 TCTGGGTAAATACCTTTTCTTCCCAAGTTTCACTATTATTTTCAATATGATCTGA 780
DB 4619 TCTGGGTAAATACCTTTTCTTCCCAAGTTTCACTATTATTTTCAATATGATCTGA 4678
QY 781 GATAGAGAAATTTTCACTAGTCTGCTGCTAAATTTTCTTATTAACCTGTTATCTTT 840
DB 4679 GATAGAGAAATTTTCACTAGTCTGCTGCTAAATTTTCTTATTAACCTGTTATCTTT 4738
QY 841 TAGGTCCTTCAGAACTCTCATTTGGTACTGAACTCAATAGGTAATCTTCACATT 900
DB 4739 TAGGTCCTTCAGAACTCTCATTTGGTACTGAACTCAATAGGTAATCTTCACATT 4798
QY 901 TATTTCTTTAGAAATAGTAATTAAGATTTTATAGCTTTTATATTTCAAGTAATGGA 960
DB 4799 TATTTCTTTAGAAATAGTAATTAAGATTTTATAGCTTTTATATTTCAAGTAATGGA 4858
QY 961 GACTATGAAAAATCCAGTTAAGTCTCTCACTGCTGTAGAGGATGATTCAGTACT 1020
DB 4859 GACTATGAAAAATCCAGTTAAGTCTCTCACTGCTGTAGAGGATGATTCAGTACT 4918
QY 1021 GTGTACTTCTGCTGCTGCAAAACAGATCACTCAACTAAGCGGCTTAAATATAG 1080
DB 4919 GTGTACTTCTGCTGCTGCAAAACAGATCACTCAACTAAGCGGCTTAAATATAG 4978
QY 1081 AACTTAAGTCTCGTAT 1098
DB 4979 AACTTAAGTCTCGTAT 4996

QY 601 TATTTCTTTTACATAGATTTTAAAGCTGAAGAGATAGTGGTCCATGATTT 660
Db 51474 TATTTCTTTTACATAGATTTTAAAGCTGAAGAGATAGTGGTCCATGATTT 51415
QY 661 ATGATCTGTGCTGGCAGGTAACCTGCTCCACAAATTTAGTTGATTTTCTTGAT 720
Db 51414 ATGATCTGTGCTGGCAGGTAACCTGCTCCACAAATTTAGTTGATTTTCTTGAT 51355
QY 721 TCTGGTAAATACCTTTTCTTCCCAAGTTTACATCTTATTTTCATATGATCTCTGA 780
Db 51354 TCTGGTAAATACCTTTTCTTCCCAAGTTTACATCTTATTTTCATATGATCTCTGA 51295
QY 781 GATPAGAAATATTTTACAGTCAGTCGCTAAATTTGTCCTATATCTGTTATCTCTT 840
Db 51294 GATPAGAAATATTTTACAGTCAGTCGCTAAATTTGTCCTATATCTGTTATCTCTT 51235
QY 841 TAGGTCTCTCCAGAAATCTCTCATTTGGTACTGAACCTCAATGGTACTTCTTCACCAAT 900
Db 51234 TAGGTCTCTCCAGAAATCTCTCATTTGGTACTGAACCTCAATGGTACTTCTTCACCAAT 51175
QY 901 TATTTCTTTAGAAATAGTAAATTTTATAGCTTTTATATTTTACCTATTTTGA 960
Db 51174 TATTTCTTTAGAAATAGTAAATTTTATAGCTTTTATATTTTACCTATTTTGA 51115
QY 961 GACTATTTGAAATCCAGTTAGCTCTCTACTGTGTGAGAGGCAATTCATGATACCT 1020
Db 51114 GACTATTTGAAATCCAGTTAGCTCTCTACTGTGTGAGAGGCAATTCATGATACCT 51055
QY 1021 GTGTACTTCTCTGTGTCTCCCAAAACAGATCACTCAATGAGCGGCTTAAATTAATG 1080
Db 51054 GTGTACTTCTCTGTGTCTCCCAAAACAGATCACTCAATGAGCGGCTTAAATTAATG 50995
QY 1081 AACTTAAGTTCTCGTAT 1098
Db 50994 AACTTAAGTTCTCGTAT 50977

RESULT 4
AF287957 168136 bp DNA linear PRI 08-JUL-2003
LOCUS Homo sapiens chromosome 8 clone CTD-2541M15 map P22-p21, complete
DEFINITION
sequence.
ACCESSION AF287957
VERSION AF287957.6 GI:32469520
KEYWORDS HTG.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 168136)
Reichwald,K., Baumgart,C., Blechschmidt,K., Dettre,M., Jahn,N.,
Wenzel,U., Polley,A., Schilhabel,M.B., Taudien,S., Wen,G. and
Rosenthal,A.
Chromosome 8 genomic sequence
2 (bases 1 to 168136)
Unpublished
3 (bases 1 to 168136)
Genome Sequencing Center Jena.
4 (bases 1 to 168136)
Submitted (14-JUL-2000) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
5 (bases 1 to 168136)
Reichwald,K. and Platzer,M.
Direct Submission
Submitted (23-AUG-2001) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
6 (bases 1 to 168136)
Lagemann,D. and Platzer,M.
Direct Submission
Submitted (29-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
7 (bases 1 to 168136)
Lagemann,D. and Platzer,M.

TITLE Direct Submission
JOURNAL Submitted (03-APR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
REFERENCE
AUTHORS Lagemann,D. and Platzer,M.
TITLE Direct Submission
JOURNAL Submitted (09-JUL-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
7 (bases 1 to 168136)
REFERENCE
AUTHORS Lagemann,D. and Platzer,M.
TITLE Direct Submission
JOURNAL Submitted (08-JUL-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
On Jul 8, 2003 this sequence version replaced gi:21717104.
COMMENT
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: <http://genome.imb-jena.de/>
Contact: gecj-submit@genome.imb-jena.de
Project Information
Center project name: H351
Center clone name: CTD-2541M15
Summary statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167707 bases at least Q40
Consensus quality: 168123 bases at least Q30
Consensus quality: 168136 bases at least Q20
Quality coverage: 6.18x

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one pUC18 subclone, and the
assembly was confirmed by restriction digest.

Neighboring sequence information:
This clone is overlapped by GS1-24F4, GS1-24F4.

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

Location/Qualifiers
1. 168136
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="p22-p21"
/clone="CTD-2541M15"
165266..168136
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="GS1-24F4"
/note="overlapping clone, overlapping clone"
6418
/note="low quality region"
7442..7645
/note="single stranded/single chemistry region"
28243..28507
/note="single stranded/single chemistry region"
28499

misc_feature /note="low quality region"
28501
misc_feature /note="low quality region"
34532..34843
misc_feature /note="single stranded/single chemistry region"
35987..36115
misc_feature /note="single stranded/single chemistry region"
37530
misc_feature /note="low quality region"
38752..38820
misc_feature /note="single stranded/single chemistry region"
38995..39017
misc_feature /note="single stranded/single chemistry region"
44702..44825
misc_feature /note="single stranded/single chemistry region"
45312..45406
misc_feature /note="single stranded/single chemistry region"
47601..47995
misc_feature /note="single stranded/single chemistry region"
59027..59136
misc_feature /note="single stranded/single chemistry region"
61319
misc_feature /note="low quality region"
62590..62726
misc_feature /note="single stranded/single chemistry region"
66851..66899
misc_feature /note="single stranded/single chemistry region"
69911..70188
misc_feature /note="single stranded/single chemistry region"
75977..76021
misc_feature /note="single stranded/single chemistry region"
78279..78386
misc_feature /note="single stranded/single chemistry region"
78675
misc_feature /note="low quality region"
81011..81279
misc_feature /note="single stranded/single chemistry region"
81070..81071
misc_feature /note="low quality region"
81743..82001
misc_feature /note="single stranded/single chemistry region"
81872
misc_feature /note="low quality region"
84434..85028
misc_feature /note="pcr product sequence only"
84932..85028
misc_feature /note="single stranded/single chemistry region"
86073..86097
misc_feature /note="single stranded/single chemistry region"
86609
misc_feature /note="low quality region"
89636..89783
misc_feature /note="single stranded/single chemistry region"
90439..90940
misc_feature /note="single stranded/single chemistry region"
95593..95693
misc_feature /note="single stranded/single chemistry region"
98827..98848
misc_feature /note="single stranded/single chemistry region"
98860
misc_feature /note="low quality region"
99495..99715
misc_feature /note="single stranded/single chemistry region"
104816..105145
misc_feature /note="single stranded/single chemistry region"
105357..105404
misc_feature /note="single stranded/single chemistry region"
105600..106813
misc_feature /note="pcr product sequence only"
112247..112576
misc_feature /note="single stranded/single chemistry region"
117025..117051
misc_feature /note="single stranded/single chemistry region"

misc_feature 118344..118414
/note="single stranded/single chemistry region"
misc_feature 141144..141340
/note="single stranded/single chemistry region"
misc_feature 141695..141700
/note="single stranded/single chemistry region"
misc_feature 142575..142734
/note="single stranded/single chemistry region"
misc_feature 145776..145927
/note="single stranded/single chemistry region"
misc_feature 146784..146859
/note="single stranded/single chemistry region"
misc_feature 150458..150522
/note="single stranded/single chemistry region"
misc_feature 160235..161053
/note="single stranded/single chemistry region"
misc_feature 163686..163687
/note="low quality region"
misc_feature 165389..165603
/note="single stranded/single chemistry region"

ORIGIN

Query Match 90.7%; Score 996; DB 9; Length 168136;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1096; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 GTTCTGTTCCGTCATTTTATTTTATTTGATGAGGGGCAATATTTCTGT 60
DB 116065 GTTCTGTTCCGTCATTTTATTTTATTTGATGAGGGGCAATATTTCTGT 116124
QY 61 TTAGTACAGAGCAGGATTTTGAATTTATTTAGGCTCTTTTCGAGCTGGGTTCC 120
DB 116125 TTAGTACAGAGCAGGATTTTGAATTTATTTAGGCTCTTTTCGAGCTGGGTTCC 116184
QY 121 TGTGTACCAAGCTACCTTCAATTTTATTTTATTTGTTTCTGTTAAGATTAAATCAATAG 180
DB 116185 TGTGTACCAAGCTACCTTCAATTTTATTTTATTTGTTTCTGTTAAGATTAAATCAATAG 116244
QY 181 AGGAATTAATGCTATCTTGAACATTAAGACCAAGAAAGAAATTAATGATGCTC 240
DB 116245 AGGAATTAATGCTATCTTGAACATTAAGACCAAGAAAGAAATTAATGATGCTC 116304
QY 241 TGTACCTTATTTTTCCTGTCGATTTGATCACTTAATCTTGTCACTGAGATGTTTGA 300
DB 116305 TGTACCTTATTTTTCCTGTCGATTTGATCACTTAATCTTGTCACTGAGATGTTTGA 116364
QY 301 TTAATAATTTTATGCTGTTTCTGTTTGTATGAGACATCTTTTCTTGAATGTT 360
DB 116365 TTAATAATTTTATGCTGTTTCTGTTTGTATGAGACATCTTTTCTTGAATGTT 116424
QY 361 TTTATCAGCTTTCGTTGCAAGGCTAGTGAATTTCTGTTCTGATTAAGATTGTT 420
DB 116425 TTTATCAGCTTTCGTTGCAAGGCTAGTGAATTTCTGTTCTGATTAAGATTGTT 116484
QY 421 GACATATTTGAGAGGATTTTATGTAATTTAAGAGTTATTAAGTTTAAATAAAGT 480
DB 116485 GACATATTTGAGAGGATTTTATGTAATTTAAGAGTTATTAAGTTTAAATAAAGT 116544
QY 481 TTATTAATTTATATATTAAGAAGCATTTTAAATTAATTTTAAATGAATTT 540
DB 116545 TTATTAATTTATATATTAAGAAGCATTTTAAATTAATTTTAAATGAATTT 116604
QY 541 TTACACCTTTCAACTCTAGTTTAAATAATAGGTTTCAAGTATTTCTGCAAGAA 600
DB 116605 TTACACCTTTCAACTCTAGTTTAAATAATAGGTTTCAAGTATTTCTGCAAGAA 116664
QY 601 TATTTTCTTTTACATTAATTTTAAAGTGAAGAGATAGTAGTCCATGAGATTT 660
DB 116665 TATTTTCTTTTACATTAATTTTAAAGTGAAGAGATAGTAGTCCATGAGATTT 116724
QY 661 ATGATCTGTGCTGGAGTAACCTGCTTCAACAAATTTAGTTGATTTTCTTGAT 720
DB 116725 ATGATCTGTGCTGGAGTAACCTGCTTCAACAAATTTAGTTGATTTTCTTGAT 116784

TITLE
JOURNAL
COMMENT

Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.

Submitted (06-AUG-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 6, 2002 this sequence version replaced g1:15290844.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: MIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu

Project Information
Project name: L12159
Center project name: 782_N1
Summary Statistics
Sequencing vector: Plasmid; n/a; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.960731
Consensus quality: 206063 bases at least Q40
Consensus quality: 207712 bases at least Q30
Consensus quality: 208309 bases at least Q20
Insert size: 205000; agarose-fp
Insert size: 208818; sum-of-contigs
Quality coverage: 8.8 in Q20 bases; agarose-fp
Quality coverage: 8.7 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently consists of 10 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 74411: contig of 74411 bp in length
* 74412 74511: gap of 100 bp
* 74512 74749: contig of 238 bp in length
* 74750 74849: gap of 100 bp
* 74850 76857: contig of 2008 bp in length
* 76858 76957: gap of 100 bp
* 76958 80208: contig of 3251 bp in length
* 80209 80308: gap of 100 bp
* 80309 88877: contig of 8569 bp in length
* 88878 88977: gap of 100 bp
* 88978 104129: contig of 15152 bp in length
* 104130 104229: gap of 100 bp
* 104230 120104: contig of 15875 bp in length
* 120105 120204: gap of 100 bp
* 120205 152189: contig of 31985 bp in length
* 152190 152289: gap of 100 bp
* 152290 208329: contig of 56040 bp in length
* 208330 208429: gap of 100 bp
* 208430 209718: contig of 1289 bp in length.

FEATURES

SOURCE

1. 209718
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-782N1"
/clone_lib="RP11 Human Male BAC"
1. 74411
/note="assembly_fragment"
clone_end:SP6
vector_side:left
74512_74749
/note="assembly_fragment"
74850_76857
/note="assembly_fragment"

misc_feature 76958..80208
/note="assembly_fragment"
misc_feature 80309..88877
/note="assembly_fragment"
misc_feature 88978..104129
/note="assembly_fragment"
misc_feature 104230..120104
/note="assembly_fragment"
misc_feature 120205..152189
/note="assembly_fragment"
misc_feature 152290..208329
/note="assembly_fragment"
misc_feature 208430..209718
/note="assembly_fragment"
clone_end:T7
vector_side:right

ORIGIN

Query Match 90.7%; Score 996; DB 2; Length 209718;
Best Local Similarity 99.8%; Pred. No. 0;
Matches 1096; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 GTTCTCTGTCCTCCGATTTTATTTATGTATGAGGAGGACAAATATTTCTGT 60
Db 130520 GTTCTCTGTCCTCCGATTTTATTTATGTATGAGGAGGACAAATATTTCTGT 130579
61 TTAGTAACAGACGAGGATTTGAATTTATAGGCTTTTCTGCAAGTCTGGTTCC 120
Db 130580 TTAGTAACAGACGAGGATTTGAATTTATAGGCTTTTCTGCAAGTCTGGTTCC 130639
121 TGTGTACCAAGGCTACCTTCAATTTTATTTATTTCTGTTAAGTTAATCAATAG 180
Db 130640 TGTGTACCAAGGCTACCTTCAATTTTATTTATTTCTGTTAAGTTAATCAATAG 130699
181 AGGAATTAATGCAATCTTCAATTAAGACCAAGGAAAGAAATTAATGATGTC 240
Db 130700 AGGAATTAATGCAATCTTCAATTAAGACCAAGGAAAGAAATTAATGATGTC 130759
241 TGTCACTTATTTTACCTGCTGATTTGTACATTAATTTGTCACTGATGATTTGA 300
Db 130760 TGTCACTTATTTTACCTGCTGATTTGTACATTAATTTGTCACTGATGATTTGA 130819
301 TTTAAATTTTACCTGCTTCTGTTTGTGTTGTTAGACACCTTTTCTTGAATGTT 360
Db 130820 TTTAAATTTTACCTGCTTCTGTTTGTGTTGTTAGACACCTTTTCTTGAATGTT 130879
361 TTTATCACTTCTGCTGCAAGGCTAGATGATCTCTGTTGTTAATGATTTGT 420
Db 130880 TTTATCACTTCTGCTGCAAGGCTAGATGATCTCTGTTGTTAATGATTTGT 130939
421 GACTCACTTCTGAAGGAGTTTATGATTTAAGAGTTAATGATTTTAAATAAAGT 480
Db 130940 GACTCACTTCTGAAGGAGTTTATGATTTAAGAGTTAATGATTTTAAATAAAGT 130999
481 TTATTAATTTAATATTAATTAAGAGCAATTTAAATTAATTTTAAATGACATT 540
Db 131000 TTATTAATTTAATATTAATTAAGAGCAATTTAAATTAATTTTAAATGACATT 131059
541 TTACACCTTTCAACCTTGAAGTTTAAATTAAGAGTTTCAAGATTTCTGCAAGAA 600
Db 131060 TTACACCTTTCAACCTTGAAGTTTAAATTAAGAGTTTCAAGATTTCTGCAAGAA 131119
601 TATTTCTTTTACATAGATTTTAAAGTGAAGAGATAGTAGTCCATGAGATT 660
Db 131120 TATTTCTTTTACATAGATTTTAAAGTGAAGAGATAGTAGTCCATGAGATT 131179
661 ATGATCTGCTGGCAGGTAACCTGCTTCAACAAATTTAGTTGATTTTCTTGAT 720
Db 131180 ATGATCTGCTGGCAGGTAACCTGCTTCAACAAATTTAGTTGATTTTCTTGAT 131239
721 TCTGGTAAATACCTTTTCTCCCAAGTTCACTACTTATTTTCAATGATCTCTGA 780
Db 131240 TCTGGTAAATACCTTTTCTCCCAAGTTCACTACTTATTTTCAATGATCTCTGA 131299

QY 781 GATAGAGAAATATTTCAGTCAGTCGCTAAATTTGTTCTTATTAACHTGTTATTCCTT 840
|||
Db 131300 GATAGAGAAATATTTCAGTCAGTCGCTAAATTTGTTCTTATTAACHTGTTATTCCTT 131359
|||
QY 841 TAGGTCCTCCAGAAATCTCTCATTTGTAAGTAACTCAATGGGACTTTCTTCACCAT 900
131360 TAGGTCCTCCAGAAATCTCTCATTTGTAAGTAACTCAATGGGACTTTCTTCACCAT 131419
|||
QY 901 TATTTCTTAAATTAAGTAATTAAGATTTTATTAAGCTTTTATTTATTTACAGTAATTTGA 960
131420 TATTTCTTAAATTAAGTAATTAAGATTTTATTAAGCTTTTATTTATTTACAGTAATTTGA 131479
|||
QY 961 GACTATTGAAATCAAGTTAAGTCTCTCTACTGTTGAGAGGACTGATTCAGTACTCT 1020
131480 GACTATTGAAATCAAGTTAAGTCTCTCTACTGTTGAGAGGACTGATTCAGTACTCT 131539
|||
QY 1021 GTGTTACTTCTCTGTGTCGCAAAAAGATGACCTCAAACTAAGGGCTTAAATTAATAG 1080
131540 GTGTTACTTCTCTGTGTCGCAAAAAGATGACCTCAAACTAAGGGCTTAAATTAATAG 131599
|||
QY 1081 AACTTAAGTTCTCGTGAT 1098
131600 AACTTAAGTTCTCGTGAT 131617
|||

RESULT 7

AC009632

LOCUS Homo sapiens chromosome 8 clone RP11-397K22 map 8, *** SEQUENCING
DEFINITION IN PROGRESS ***, 48 unordered pieces.

AC009632

AC009632.3 GI:8954116
HTG: HTGS, PHASE1

KEYWORDS Homo sapiens (human)

SOURCE

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 185834)
Britten, B., Linton, L., Nussbaum, C. and Lander, E.

TITLE

Unpublished
2 (bases 1 to 185834)

REFERENCE

Britten, B., Linton, L., Nussbaum, C., Lander, E., Allen, N., Anderson, M.,
Baker, J., Baldwin, J., Barna, N., Beckerly, R., Benn, J., Brown, A.,
Cacile, A., Cerny, J., Colangelo, M., Collins, S., Collamore, A.,
Cooke, P., Dearellano, K., Depayre, B., Devon, K., Dewar, K.,
Domelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,
Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
Karats, A., Lehoczy, J., Lieu, C., Locke, K., Macdonald, P.,
Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
Meldrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
Naylor, J., Nijloff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
Teefaye, S., Tornella-Miller, J., Vassiliev, H., Vo, A., Wagner, A.,
Wheeler, J., Wu, X., Wymann, D., Ye, W. J. and Zody, M.

JOURNAL

Direct Submission

Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 7, 2000 this sequence version replaced gi:5932606.

All repeats were identified using RepeatMasker:

Smt, A.F.A. & Green, P. (1996-1997)

http://ftp.genome.washington.edu/RM/RepeatMasker.html

Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence_submissions@genome.wi.mit.edu

Project Information

Center project name: 397_K_22

NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 1361: contig of 1361 bp in length
1362 1461: gap of 100 bp
1462 3202: contig of 1741 bp in length
3203 3303: gap of 100 bp
3303 4619: contig of 1317 bp in length
4620 4719: gap of 100 bp
4720 5821: contig of 1102 bp in length
5822 5921: gap of 100 bp
5922 6419: contig of 498 bp in length
6420 6520: gap of 100 bp
6520 8242: contig of 1723 bp in length
8243 8343: gap of 100 bp
8343 9703: contig of 1361 bp in length
9704 9803: gap of 100 bp
9804 10936: contig of 1133 bp in length
10937 11036: gap of 100 bp
11037 12339: contig of 1303 bp in length
12340 12439: gap of 100 bp
12440 13923: contig of 1484 bp in length
13924 14023: gap of 100 bp
14024 15372: contig of 1349 bp in length
15373 15472: gap of 100 bp
15473 16957: contig of 1485 bp in length
16958 17057: gap of 100 bp
17058 19291: contig of 2234 bp in length
19292 19391: gap of 100 bp
19392 21442: contig of 2051 bp in length
21443 21542: gap of 100 bp
21543 23482: contig of 1940 bp in length
23483 23582: gap of 100 bp
23583 25557: contig of 1975 bp in length
25558 26557: gap of 100 bp
26558 28144: contig of 2487 bp in length
28145 28244: gap of 100 bp
28245 30739: contig of 2495 bp in length
30740 30839: gap of 100 bp
30840 33278: contig of 2439 bp in length
33279 33378: gap of 100 bp
33379 35479: contig of 2101 bp in length
35480 35579: gap of 100 bp
35580 38046: contig of 2467 bp in length
38047 38146: gap of 100 bp
38147 40591: contig of 2445 bp in length
40592 40691: gap of 100 bp
40692 43194: contig of 2503 bp in length
43195 43294: gap of 100 bp
43295 46358: contig of 3064 bp in length
46359 46458: gap of 100 bp
46459 49611: contig of 3153 bp in length
49612 49711: gap of 100 bp
49712 53161: contig of 3450 bp in length
53162 53261: gap of 100 bp
53262 55787: contig of 2526 bp in length
55788 55887: gap of 100 bp
55889 58727: contig of 2840 bp in length
58728 58827: gap of 100 bp
58828 61970: contig of 3143 bp in length
61971 62070: gap of 100 bp
62071 65666: contig of 3596 bp in length
65667 65766: gap of 100 bp
65767 70005: contig of 4240 bp in length
70006 70106: gap of 100 bp
70107 73614: contig of 3508 bp in length
73615 73714: gap of 100 bp

* 73715 78129: contig of 4415 bp in length
* 78130 78229: gap of 100 bp
* 78230 80902: contig of 2673 bp in length
* 80903 81002: gap of 100 bp
* 81003 84942: contig of 3340 bp in length
* 84943 85042: gap of 100 bp
* 85043 90417: contig of 5375 bp in length
* 90418 90517: gap of 100 bp
* 90518 95597: contig of 5080 bp in length
* 95598 95697: gap of 100 bp
* 95698 101500: contig of 5803 bp in length
* 101501 101600: gap of 100 bp
* 101601 108688: contig of 7088 bp in length
* 108689 108788: gap of 100 bp
* 108789 115623: contig of 6835 bp in length
* 115624 115723: gap of 100 bp
* 115724 120744: contig of 5021 bp in length
* 120745 120844: gap of 100 bp
* 120845 128016: contig of 7172 bp in length
* 128017 128116: gap of 100 bp
* 128117 133690: contig of 5574 bp in length
* 133691 133790: gap of 100 bp
* 133791 141605: contig of 7815 bp in length
* 141606 141705: gap of 100 bp
* 141706 150654: contig of 8949 bp in length
* 150655 150754: gap of 100 bp
* 150755 158712: contig of 7958 bp in length
* 158713 158812: gap of 100 bp
* 158813 168916: contig of 10104 bp in length
* 168917 169017: gap of 100 bp
* 169017 185834: contig of 16818 bp in length.

FEATURES

source

1. .185834
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-397K22"
/clone_lib="RP11 Human Male BAC"
1. .1361
/note="assembly_fragment"
1462. .3202
/note="assembly_fragment"
3303. .4619
/note="assembly_fragment"
4720. .5821
/note="assembly_fragment"
5922. .6419
/note="assembly_fragment"
clone_end:T7
vector_side:right"
6520. .8242
/note="assembly_fragment"
8343. .9703
/note="assembly_fragment"
9804. .10936
/note="assembly_fragment"
11037. .12339
/note="assembly_fragment"
12440. .13923
/note="assembly_fragment"
14024. .15372
/note="assembly_fragment"
15473. .16957
/note="assembly_fragment"
17058. .19291
/note="assembly_fragment"
19392. .21442
/note="assembly_fragment"
21543. .23482
/note="assembly_fragment"
23583. .25557
/note="assembly_fragment"

/note="assembly_fragment"
misc_feature 25658..28144
/note="assembly_fragment"
misc_feature 28245..30739
/note="assembly_fragment"
misc_feature 30840..33278
/note="assembly_fragment"
misc_feature 33379..35479
/note="assembly_fragment"

Query Match 74.0%; Score 813; DB 2; Length 185834;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 863; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

235 ATGTCGTCACCTTATTTTACCTGTGACTTGTACATTAACCTTGTGACCTGATG 294
|||||
43295 ATGTCGTCACCTTATTTTACCTGTGACTTGTGACCTTGTGACCTGATG 43354
|||||
295 TTTTGATTAATAATTTTATGCTGCTTCTGTTTGTAGACACTCTTTTCTTGA 354
|||||
43355 TTTTGATTAATAATTTTATGCTGCTTCTGTTTGTAGACACTCTTTTCTTGA 43414
|||||
355 ATGTTTATTCAGCTTCTGTTGCAAGCTAGTGAATTCCTGTTCTGATAAGT 414
|||||
43415 ATGTTTATTCAGCTTCTGTTGCAAGCTAGTGAATTCCTGTTCTGATAAGT 43474
|||||
415 ATGTCGTCACCTTATTTTACCTGTGACTTGTACATTAACCTTGTGACCTGATG 474
|||||
43475 ATGTCGTCACCTTATTTTACCTGTGACTTGTACATTAACCTTGTGACCTGATG 43534
|||||
475 AAAGTTTATTAATTAATTAATTAATTAAGGCACTTTTAAATTTTAAATG 534
|||||
43535 AAAGTTTATTAATTAATTAATTAATTAAGGCACTTTTAAATTTTAAATG 43594
|||||
535 ACATTTTACACCTTCACTGCTGCTTCACTGCTTCACTGCTTCACTGCTTCACTG 594
|||||
43595 ACATTTTACACCTTCACTGCTGCTTCACTGCTTCACTGCTTCACTGCTTCACTG 43654
|||||
595 GAAGAAATTTTCTTTTACATTAATTTTAACTGAAGAGAGTATGATGCTCATG 654
|||||
43655 GAAGAAATTTTCTTTTACATTAATTTTAACTGAAGAGAGTATGATGCTCATG 43714
|||||
655 AGATTATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 714
|||||
43715 AGATTATGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43774
|||||
715 TTGATTCGGGTAATAATCTTTTCTCCAGTTTCACTACTTTATTTTCATATGAT 774
|||||
43775 TTGATTCGGGTAATAATCTTTTCTCCAGTTTCACTACTTTATTTTCATATGAT 43834
|||||
775 CTCTGAGATGAGAAATATTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 834
|||||
43835 CTCTGAGATGAGAAATATTTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 43894
|||||
835 TCCTTTTGGCTTCCAGATCTCTCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 894
|||||
43895 TCCTTTTGGCTTCCAGATCTCTCATTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 43954
|||||
895 ACCATTTATTTCTTTAGATTAATTAATTAATTTTATTAATTTTATTTTCACTGTA 954
|||||
43955 ACCATTTATTTCTTTAGATTAATTAATTAATTTTATTAATTTTATTTTCACTGTA 44014
|||||
955 ATTTGAGACTATTTGAAATCCAGTTTAACTCTCTACGTTGTGAGAGCACTTGTCAA 1014
|||||
44015 ATTTGAGACTATTTGAAATCCAGTTTAACTCTCTACGTTGTGAGAGCACTTGTCAA 44074
|||||
1015 GTACCGTGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1074
|||||
44075 GTACCGTGTACTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 44134
|||||
1075 TAATAGACTTAAGTTCTGCTGAT 1098
|||||
44135 TAATAGACTTAAGTTCTGCTGAT 44158
|||||

RESULT 8
AC009632/C
LOCUS
DEFINITION Homo sapiens chromosome 8 clone RP11-397K22 map 8, *** SEQUENCING
IN PROGRESS ***
AC009632
AC009632.3 GI:8954116
VERSION HTGS PHASE1
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M., Baker, J., Baldwin, J., Barna, N., Beckert, R., Benn, J., Brown, A., Cacerle, A., Cerny, J., Colangelo, M., Collins, S., Collamore, A., Cooke, P., DeArnell, K., Depayre, E., Devon, K., Dewar, K., Domelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C., Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L., Karatas, A., Lehotzky, J., Lien, C., Locke, K., MacDonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J., Meltrin, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J., Naylor, J., Nilot, M., O'Connor, T., O'Donnell, P., Pavlin, B., Peterson, K., Pollara, V., Stojanovic, N., Stone, C., Subramanian, A., Stenger-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A., Testave, S., Tornella-Miller, I., Vassiliev, H., Vo, A., Wagner, A., Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
TITLE
JOURNAL
COMMENT
DIRECT SUBMISSION
Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charis Street, Cambridge, MA 02141, USA
On Jul 7, 2000 this sequence version replaced gi:5932606.
All repeats were identified using RepeatMasker:
Smt, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L2099
Center clone name: 397_K_22

* NOTE: This is a 'working draft' sequence. It currently
* consists of 48 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1
* 1361: contig of 1361 bp in length
* 1362
* 1461: gap of 100 bp
* 1462
* 3202: contig of 1741 bp in length
* 3203
* 3302: gap of 100 bp
* 3303
* 4619: contig of 1317 bp in length
* 4620
* 4719: gap of 100 bp
* 4720
* 5821: contig of 1102 bp in length
* 5822
* 5921: gap of 100 bp
* 5922
* 6419: contig of 498 bp in length
* 6420
* 6519: gap of 100 bp
* 6520
* 8342: contig of 1723 bp in length
* 8343
* 8423: gap of 100 bp
* 8424
* 8343: contig of 1361 bp in length
* 8343
* 9703: gap of 100 bp
* 9704
* 10936: contig of 1133 bp in length
* 9804

* 10937
* 11037
* 12339: contig of 1303 bp in length
* 12340
* 12439: gap of 100 bp
* 12440
* 13923: contig of 1484 bp in length
* 13924
* 14023: gap of 100 bp
* 14024
* 15372: contig of 1349 bp in length
* 15373
* 15472: gap of 100 bp
* 15473
* 16957: contig of 1485 bp in length
* 16958
* 17057: gap of 100 bp
* 17058
* 19291: contig of 2234 bp in length
* 19292
* 19391: gap of 100 bp
* 19392
* 21442: contig of 2051 bp in length
* 21443
* 21543
* 23482: contig of 1940 bp in length
* 23483
* 23582: gap of 100 bp
* 23583
* 25557: contig of 1975 bp in length
* 25558
* 25657: gap of 100 bp
* 25658
* 28144: contig of 2487 bp in length
* 28145
* 30739: contig of 2495 bp in length
* 30740
* 30839: gap of 100 bp
* 30840
* 33278: contig of 2439 bp in length
* 33279
* 33378: gap of 100 bp
* 33379
* 35479: contig of 2101 bp in length
* 35480
* 35579: gap of 100 bp
* 35580
* 38046: contig of 2467 bp in length
* 38047
* 38146: gap of 100 bp
* 38147
* 40591: contig of 2445 bp in length
* 40592
* 40691: gap of 100 bp
* 40692
* 43194: contig of 2503 bp in length
* 43195
* 43294: gap of 100 bp
* 43295
* 46358: contig of 3064 bp in length
* 46359
* 46458: gap of 100 bp
* 46459
* 46611: contig of 3153 bp in length
* 46612
* 49711: gap of 100 bp
* 49712
* 53161: contig of 3450 bp in length
* 53162
* 53261: gap of 100 bp
* 53262
* 55787: contig of 2526 bp in length
* 55788
* 55887: gap of 100 bp
* 55888
* 58727: contig of 2840 bp in length
* 58728
* 58729: gap of 100 bp
* 58728
* 61970: contig of 3143 bp in length
* 61971
* 62071: gap of 100 bp
* 62071
* 65666: contig of 3596 bp in length
* 65667
* 65765: gap of 100 bp
* 65767
* 70006: contig of 4240 bp in length
* 70007
* 70106: gap of 100 bp
* 70107
* 73114: contig of 3508 bp in length
* 73115
* 73116: gap of 100 bp
* 73115
* 78129: contig of 4415 bp in length
* 78130
* 78229: gap of 100 bp
* 78230
* 80902: contig of 2673 bp in length
* 80903
* 81002: gap of 100 bp
* 81003
* 84942: contig of 3940 bp in length
* 84943
* 85042: gap of 100 bp
* 85043
* 90417: contig of 5375 bp in length
* 90418
* 90517: gap of 100 bp
* 90518
* 95597: contig of 5080 bp in length
* 95597
* 95697: gap of 100 bp
* 95698
* 101500: contig of 5803 bp in length
* 101501
* 101600: gap of 100 bp
* 101601
* 106888: contig of 7088 bp in length
* 106889
* 108788: gap of 100 bp
* 108789
* 115623: contig of 6835 bp in length
* 115624
* 115723: gap of 100 bp
* 115724
* 120744: contig of 5021 bp in length
* 120745
* 120844: gap of 100 bp
* 120845
* 128016: contig of 7172 bp in length
* 128017
* 128116: gap of 100 bp
* 128117
* 133690: contig of 5574 bp in length
* 133691
* 133790: gap of 100 bp
* 133791
* 141605: contig of 7815 bp in length
* 141705: gap of 100 bp
* 141606

FEATURES
source

1. 15834
Location/Qualifiers

1. 15834
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"
/clone="RP11-397K22"
/clone_1ib="RP11-11 Human Male BAC"
1. 1361
/note="assembly_fragment"
1462. 3202
/note="assembly_fragment"
3303. 4619
/note="assembly_fragment"
4720. 5821
/note="assembly_fragment"
5922. 6419
/note="assembly_fragment"
clone_end:17
vector_side:right"
6520. 8242
/note="assembly_fragment"
8343. 9703
/note="assembly_fragment"
9804. 10936
/note="assembly_fragment"
11037. 12339
/note="assembly_fragment"
12440. 13923
/note="assembly_fragment"
14024. 15372
/note="assembly_fragment"
15473. 16957
/note="assembly_fragment"
17058. 19291
/note="assembly_fragment"
19392. 21442
/note="assembly_fragment"
21543. 23482
/note="assembly_fragment"
23583. 25557
/note="assembly_fragment"
25658. 28144
/note="assembly_fragment"
28245. 30739
/note="assembly_fragment"
30840. 33278
/note="assembly_fragment"
33379. 35479
/note="assembly_fragment"

Query Match 10.4%; Score 114; DB 2; Length 185834;
Best Local Similarity 99.4%; Pred. No. 4e-42;
Matches 164; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GTTCTGTCCTCCATTTTATTTATGTATGAGGGGACAAATATTTCTGT 60
DB 101765 GTTCTGTCCTCCATTTTATTTATGTATGAGGGGACAAATATTTCTGT 101706
QY 61 TTAGTAACAGACGAGGATTTTGAATTTATAGGCTTTTTCGACGCTGGGTTCC 120
DB 101705 TTAGTAACAGACGAGGATTTTGAATTTATAGGCTTTTTCGACATCTGGGTTCC 101646
QY 121 TGTGTACACAAAGCTACTTCAATTTTATTTGTTCTGTGA 165

DB 101645 TGTGTACACAAAGCTACTTCAATTTTATTTGTTCTGTGA 101601

RESULT 9
LOCUS AX345694 19087 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 765 from Patent WO0200928.
ACCESSION AX345694
VERSION AX345694.1 GI:18493580

KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
artificial sequences.

REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 765 03-JAN-2002;
Epigenomics AG (DE)
LOCATION/Qualifiers

FEATURES
source
1. 19087
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 2.3%; Score 25; DB 6; Length 19087;
Best Local Similarity 100.0%; Pred. No. 1.3;
Matches 25; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 509 TTTTAAATTAATTTTATTTTAAAT 533
DB 14077 TTTTAAATTAATTTTATTTTAAAT 14101

RESULT 10
LOCUS AX345679 6782 bp DNA linear PAT 01-FEB-2002
DEFINITION Sequence 750 from Patent WO0200928.
ACCESSION AX345679
VERSION AX345679.1 GI:18493565

KEYWORDS
SOURCE
ORGANISM
synthetic construct
synthetic construct
artificial sequences.

REFERENCE
1
AUTHORS Olek, A., Piepenbrock, C. and Berlin, K.
TITLE Diagnosis of diseases associated with the immune system
JOURNAL Patent: WO 0200928-A 750 03-JAN-2002;
Epigenomics AG (DE)
LOCATION/Qualifiers

FEATURES
source
1. 6782
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN

Query Match 2.2%; Score 24; DB 6; Length 6782;
Best Local Similarity 100.0%; Pred. No. 4.3;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 ATTTTATTTTATTTATGTATGAGG 39
DB 2770 ATTTTATTTTATTTATGTATGAGG 2793

RESULT 11
LOCUS AC134447 179514 bp DNA linear ROD 12-DEC-2003
DEFINITION Mus musculus chromosome 18 clone RP24-166D1, complete sequence.
ACCESSION AC134447
VERSION AC134447.2 GI:39752882

KEYWORDS
SOURCE HTG.
ORGANISM Mus musculus (house mouse)

REFERENCE
AUTHORS Bukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
TITLE 1 (bases 1 to 179514)
JOURNAL Wilson, R.K.
The sequence of Mus musculus clone

REFERENCE
AUTHORS 2 (bases 1 to 179514)
TITLE McPherson, J.D. and Waterston, R.H.
JOURNAL Direct Submission
Submitted (26-SEP-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
3 (bases 1 to 179514)

REFERENCE
AUTHORS Wilson, R.K.
TITLE Direct Submission
JOURNAL Submitted (12-DEC-2003) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA
On Dec 12, 2003 this sequence version replaced gi:23322797.

COMMENT

----- Genome Center -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu
Contact: submissions@wustl.edu
Project Information
Center project name: M_BB0186D01

FEATURES
source
Location/Qualifiers
1. 179514
/organism="Mus musculus"
/mol_type="genomic DNA"
/db_xref="taxon:10090"
/chromosome="18"
/clone="RP24-186D1"

ORIGIN

Query Match 2.2% Score 24; DB 10; Length 179514;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 TTTATTAATTATATATTAAG 503
Db 141070 TTTATTAATTATATATTAAG 141047

RESULT 12
AC135128/c
LOCUS AC135128 264864 bp DNA linear HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-249P17, *** SEQUENCING IN PROGRESS
***, 2 unordered pieces.
AC135128
AC135128.2 GI:25138667
HTG: HTGS PHASE1; HTGS DRAFT; HTGS_ENRICHED.
KEYWORDS Rattus norvegicus (Norway rat)
SOURCE Rattus norvegicus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
1 (bases 1 to 264864)
Muzny, D., Maric, M., Metzger, M., Lee, S., Abramson, S., Adams, C., Alder, J., Allen, C., Allen, H., Albrooks, S., Amin, A., Angiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, W., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Behnke, F., Bialwalo, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhey, C., Burch, P., Butrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M., Davis, C., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, O., Denson, S., Deremo, C., Ding, Y., Dinh, H., Divya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K.,

Egan, A., Escotto, M., Eugene, C., Evans, C.A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Fraser, C.M., Gabisi, A., Ganta, R., Garcia, A., Garner, T., Garza, M., Gebregeorgis, E., Geer, K., Gill, R., Grady, M., Guerra, W., Guevara, W., Gunaratne, P., Haaland, M., Hamill, C., Hamilton, C., Hamilton, K., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S.L., Hodgson, A., Hogues, M., Hollins, B., Howell, S., Hulvik, S., Hume, J., Idlebird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowis, C., Kraft, C.L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenshewa, L., Loulsged, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindartine, M., Mahmood, M., Malloy, K., Mangum, A., Mangum, B., Mapa, P., Martin, K., Martin, R., Martinez, E., Mawhney, S., McLeod, M.P., McNeill, T.Z., Meenen, E., Miosavljevic, A., Miner, G., Minda, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwokedeme, O., Okunolu, G., Olarnpunsagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Poldexter, A., Popovic, D., Primus, E., Pu, L., Puzo, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., J., Sanders, W., Savary, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sison, I., Sitter, C.D., Smaje, D., Sneed, J., Sodergren, E., Song, X.-Z., Sorelle, R., Sosa, J., Steimle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villasana, D., Waldron, L., Walker, B., Wang, J., Wang, O., Wang, S., Warren, J., Warren, R., Wei, X., White, P., Williams, G., Willson, R., Wleczky, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Zhu, D., von Weinstock, G., and Gibbs, R.A.

TITLE
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 264864)
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (07-OCT-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
3 (bases 1 to 264864)
Rat Genome Sequencing Consortium.
REFERENCE
AUTHORS Rat Genome Sequencing Consortium.
TITLE Direct Submission
JOURNAL Submitted (20-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
On Nov 20, 2002 this sequence version replaced gi:23507717.
The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center -----
Center: Baylor College of Medicine
Center code: BCM
Web site: <http://www.hgsc.bcm.tmc.edu/>
Contact: hgsc-help@bcm.tmc.edu
----- Project Information -----
Center project name: KCZU
Center clone name: CH230-249P17
----- Summary Statistics -----

Assembly program: Phrap: version 0.990329
Consensus quality: 239879 bases at least Q40
Consensus quality: 242628 bases at least Q20
Consensus quality: 244265 bases at least Q20
Estimated insert size: 242071; sum-of-contigs estimation
Quality coverage: 6x in Q20 bases; sum-of-contigs estimation

NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 2 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.
1 263304: contig of 263304 bp in length
* 263305 263404: gap of unknown length
* 263405 264864: contig of 1460 bp in length.
Location/Qualifiers

FEATURES
source
1..264864
/organism="Rattus norvegicus"
/mol_type="genomic DNA"
/db_xref="taxon:10116"
/clone="CH230-249F17"
1..1215
/note="wgs_end_extension
clone_end:Sp6"
4648..5394
/note="clone boundary
clone_end:Sp6
site:"
end_sequence:BB213261"
100740..102557
/note="wgs_contig"
complement(259844..260595)
/note="clone boundary
clone_end:T7
site:
end_sequence:BB213258"
261463..263304
/note="wgs_end_extension
clone_end:T7"

ORIGIN

Query Match 2.2%; Score 24; DB 2; Length 264864;
Best Local Similarity 100.0%; Pred. No. 2.6;
Matches 24; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 599 AATATTTCTTTACATAGATTT 622
DB 195518 AATATTTCTTTACATAGATTT 195495

RESULT 13
AR290033/c 47 bp DNA linear PAT 12-JUN-2003
LOCUS AR290033 Sequence 1768 from patent US 6537751.
DEFINITION AR290033
ACCESSION AR290033
VERSION AR290033.1 GI:31677317
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE Unclassified.
1 (bases 1 to 47)
AUTHORS Cohen,D., Chumakov,I. and Blumenfeld,M.
TITLE Biallelic markers for use in constructing a high density
disequilibrium map of the human genome
JOURNAL Patent: US 6537751-A 1768 25-MAR-2003;
FEATURES Location/Qualifiers
source 1..47
/organism="unknown"

ORIGIN /mol_type="genomic DNA"

Query Match 2.1%; Score 23; DB 6; Length 47;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 531 AATGACATTTTACACCTTTCAA 553
DB 47 AATGACATTTTACACCTTTCAA 25
RESULT 14
AX114392 47 bp DNA linear PAT 11-MAY-2001
LOCUS AX114392 Sequence 61 from Patent WO0129257.
DEFINITION AX114392
ACCESSION AX114392
VERSION AX114392.1 GI:14031356
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 Schork,N. and Skierczynski,B.
AUTHORS Methods of genetic cluster analysis and use thereof
TITLE Patent: WO 0129257-A 61 26-APR-2001;
JOURNAL GENSET (PR)
FEATURES Location/Qualifiers
source 1..47
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
variation 24
/note="99-622-95 : polymorphic base A or G"

ORIGIN
Query Match 2.1%; Score 23; DB 6; Length 47;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 531 AATGACATTTTACACCTTTCAA 553
DB 47 AATGACATTTTACACCTTTCAA 25

RESULT 15
AX281468/c 6070 bp DNA linear PAT 02-NOV-2001
LOCUS AX281468 Sequence 132 from Patent WO0177376.
DEFINITION AX281468
ACCESSION AX281468
VERSION AX281468.1 GI:16608723
KEYWORDS
SOURCE synthetic construct
ORGANISM synthetic construct
REFERENCE Olek,A., Piepenbrock,C. and Berlin,K.
AUTHORS Diagnosis of diseases associated with metastasis
TITLE Patent: WO 0177376-A 132 18-OCT-2001;
JOURNAL Epigenomics AG (DE)
FEATURES Location/Qualifiers
source 1..6070
/organism="synthetic construct"
/mol_type="unassigned DNA"
/db_xref="taxon:32630"
/note="chemically treated genomic DNA (Homo sapiens)"

ORIGIN
Query Match 2.1%; Score 23; DB 6; Length 6070;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Thu May 27 08:41:18 2004

us-09-901-484a-179_3899_4996.01lg.rge

Page 17

QY	480	TTTATTAATTATATATTTAA	502
Db	2617	TTTATTAATTATATATTTAA	2795

Search completed: May 26, 2004, 10:29:41
Job time : 4181.21 secs

THIS PAGE LEFT BLANK

Db	1021	ATATCTTGTTGGGATTGTTCTTCAAGCTATGATATAACATGTAATTATCATTAATCTGTG	1080
Qy	1081	ATGGGGAAAAAACACGGAACCTTAATTCTGAAAACACCTTGATGCAAGACGGGACGAGAGG	1140
Db	1081	ATGGGGAAAAAACACGGAACCTTAATTCTGAAAACACCTTGATGCAAGACGGGACGAGAGG	1140
Qy	1141	GGCTGCTGCGGACATCCAGAGCGGAGGCTGAGGAGGCGGCTGCCCTTGCAAAAGCACTGGCA	1200
Db	1141	GGCTGCTGCGGACATCCAGAGCGGAGGCTGAGGAGGCGGCTGCCCTTGCAAAAGCACTGGCA	1200
Qy	1201	GTGAGCAGATGGGGACACTCGAGCTGCCCCCGACACTTGAGCGGACTGACCCTACACCTGG	1260
Db	1201	GTGAGCAGATGGGGACACTCGAGCTGCCCCCGACACTTGAGCGGACTGACCCTACACCTGG	1260
Qy	1261	GCCGAGGTGCTCTGCAAGATTAGAATTCCGATPAAAGTTAACACCACATTTCTCACTGTCTC	1320
Db	1261	GCCGAGGTGCTCTGCAAGATTAGAATTCCGATPAAAGTTAACACCACATTTCTCACTGTCTC	1320
Qy	1321	TAAATTGTGTGATCCCGGCGCCCGAGGGGCTTGTGACAGCAGGTGCGGCTTCCAGGACAC	1380
Db	1321	TAAATTGTGTGATCCCGGCGCCCGAGGGGCTTGTGAGCAGCAGGTGCGGCTTCCAGGACAC	1380
Qy	1381	TCCAGCGCACCTTTAAACCTGACCGGCGGCAAGTCGCGGCGGAGGAGGAGCAAAAGAGGC	1440
Db	1381	TCCAGCGCACCTTTAAACCTGACCGGCGGCAAGTCGCGGCGGAGGAGGAGCAAAAGAGGC	1440
Qy	1441	ACCAGGACCCCTCTCCGGCGACGACCCACCTTCAACCCAGTTCCGTCAGTCGCCACACCT	1500
Db	1441	ACCAGGACCCCTCTCCGGCGACGACCCACCTTCAACCCAGTTCCGTCAGTCGCCACACCT	1500
Qy	1501	CCCTTCCCGCGCTCCGCAAGCCCGGCCCAGTCGGGGAGCATGCGCAGTGGCCGAGCCGGGTT	1560
Db	1501	CCCTTCCCGCGCTCCGCAAGCCCGGCCCAGTCGGGGAGCATGCGCAGTGGCCGAGCCGGGTT	1560
Qy	1561	GCCGCGCGCAAGAGGTGATCTGCACTGTCCGCGCCCAACCAACCAATCAAGAGA	1620
Db	1561	GCCGCGCGCAAGAGGTGATCTGCACTGTCCGCGCCCAACCAACCAATCAAGAGA	1620
Qy	1621	CGTGTATTGCGCGCGAGGTGAACTATAGCGAACCGGCGAACCAATCAGAAAGCGCGTGTGT	1680
Db	1621	CGTGTATTGCGCGCGAGGTGAACTATAGCGAACCGGCGAACCAATCAGAAAGCGCGTGTGT	1680
Qy	1681	TGCCGCGGAGCCCCCTGCCCCGCGGACGAGGGAGTGTGCATGGGTGAGGGGTATGAGGGGTGT	1740
Db	1681	TGCCGCGGAGCCCCCTGCCCCGCGGACGAGGGAGTGTGCATGGGTGAGGGGTATGAGGGGTGT	1740
Qy	1741	GAGCATCCCTBAGCATCATGATCCCGGAGGGCGCGGGATTCCTTGTGCGCGCGGAG	1800
Db	1741	GAGCATCCCTBAGCATCATGATCCCGGAGGGCGCGGGATTCCTTGTGCGCGCGGAG	1800
Qy	1801	CGGCGCACAGCAGCCCCGCACTGCGCTTACCGBGCCCCGAGCGGCGGCGCCATGCGGC	1860
Db	1801	CGGCGCACAGCAGCCCCGCACTGCGCTTACCGBGCCCCGAGCGGCGGCGCCATGCGGC	1860
Qy	1861	TGGGGGCGGAGGCTTGAGAGCGGGTGTGGCGGCGCGCGCGGCCCCGAGCGGTGATTGG	1920
Db	1861	TGGGGGCGGAGGCTTGAGAGCGGGTGTGGCGGCGCGCGGCCCCGAGCGGTGATTGG	1920
Qy	1921	CCGCTGTGTGGCCGCGACTGAGGCCCCGGGAGCGGGGAGCGCAGAGCGGAGTCCGCT	1980
Db	1921	CCGCTGTGTGGCCGCGACTGAGGCCCCGGGAGCGGGGAGCGCAGAGCGGAGTCCGCT	1980
Qy	1981	GCGCGCAGCTGAGAAAGATGCTGCTGTCCCTGTGTGCTCACACAGTACTCAATGCGCTAAC	2040
Db	1981	GCGCGCAGCTGAGAAAGATGCTGCTGTCCCTGTGTGCTCACACAGTACTCAATGCGCTAAC	2040
Qy	2041	TGCTGTCCAGAGGTGCTGTCTCTGTGGGCAAGGCGCCACCTACGTGTGTGGCTGTGGGGGGCT	2100
Db	2041	TGCTGTCCAGAGGTGCTGTCTCTGTGGGCAAGGCGCCACCTACGTGTGTGTGGCTGTGGGGGGCT	2100
Qy	2101	GGCGGCTGCTCTCCGCTTCTCTGCCCCGCGCTTTCACAAAGCGTTGAGCAACGCGGCTGT	2160
Db	2101	GGCGGCTGCTCTCCGCTTCTCTGCCCCGCGCTTTCACAAAGCGTTGAGCAACGCGGCTGT	2160

Db		2101	GAGGGGTCGTCTCGCCCTTCTTGCCCGCAGCTTCTACCAAGCGCTGAAGACGCAGCTGT	2169
Oy		2161	ACTGCGTTCTACCAAGCATATGATGCTCTTTCTTTCCGAATTACAACGGGGTCCAGGTGA	2220
Db		2161	ACTGCGTTCTACCAAGCATATGATGCTCTTTCTTTCCGAATTACAACGGGGTCCAGGTGA	2220
Oy		2221	GCGCGTCGCCGCTCCGGGTCGTGGGCGTCCAGAGTCCCGGAGGCGCGAAGCTCTCC	2280
Db		2221	GCGCGTCGCCGCTCCGGGTCGTGGGCGTCCAGAGTCCCGGAGGCGCGAAGCTCTCC	2280
Oy		2281	GCTCCCCACAGCTGCGAGGGTCAACCOCGCGGCCGCGCGAC	2324
Db		2281	GCTCCCCACAGCTGCGAGGGTCAACCOCGCGGCCGCGAC	2324
RESULT 2				
BD196564				
LOCUS				
DEFINITION		Prostatic cancer gene.	56520 bp	DNA linear PAT 17-JUL-2003
ACCESSION		BD196564		
VERSION		BD196564.1		GI:33006334
KEYWORDS		JP 2002516657-A/153.		
SOURCE		Homo sapiens (human)		
ORGANISM		Homo sapiens		
REFERENCE		Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS		Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
TITLE		Cohen, D., Blumenfeld, M., Chumakov, I. and Bouguenarel, L.		
JOURNAL		Prostatic cancer gene Patent: JP 2002516657-A 153 11-JUN-2002;		
COMMENT		GENSET		
		OS Homo sapiens (human)		
		PN JP 2002516657-A/153		
		PD 11-JUN-2002		
		PF 22-DEC-1998 JP 2000525562		
		PR 22-DEC-1997 US 08/996306, 09-SEP-1998 US 60/099658 PI		
		DANIEL COHEN, MARTA BLUMENFELD, ILIYA CHUMAKOV, LYDIE BOUGUENAREL PC		
		CI2N15/09, CI2N15/09, A01K67/027, C07K14/47, C07K16/18, CI2N1/15, PC		
		CI2N1/19,		
		PC CI2N1/21, CI2N5/10, CI2N5/10, CI2P21/08, CI2O1/68, GO1N33/50 PC		
		PC CI2N15/00, CI2N5/00,		
		PC CI2N5/00, CI2N15/00		
		CC exon1		
		CC exon2		
		CC exon3		
		CC exon4		
		CC exon5		
		CC exon6		
		CC exon7		
		CC exon8		
		CC AATAAA		
		CC upstream amplification primer 5-63		
		CC downstream amplification primer 5-63 , complement CC		
		CC upstream amplification primer 99-622		
		CC upstream amplification primer 99-622 , complement CC		
		CC downstream amplification primer 99-621		
		CC upstream amplification primer 99-621 , complement CC		
		CC downstream amplification primer 99-619		
		CC upstream amplification primer 99-619 , complement CC		
		CC upstream amplification primer 4-76		
		CC downstream amplification primer 4-76 , complement CC		
		CC upstream amplification primer 4-77		
		CC downstream amplification primer 4-77 , complement CC		
		CC upstream amplification primer 4-71		
		CC downstream amplification primer 4-71 , complement CC		
		CC upstream amplification primer 4-72		
		CC downstream amplification primer 4-72 , complement CC		
		CC upstream amplification primer 4-73		
		CC downstream amplification primer 4-73 , complement CC		
		CC upstream amplification primer 99-610		
		CC downstream amplification primer 99-610 , complement CC		
		CC upstream amplification primer 99-609		
		CC upstream amplification primer 99-609 , complement CC		

[illegible]

Query Match

100.0%; Score 2324; DB 6; Length 56520;

Best Local Similarity 100.0%; Pred. No. 0;
Matches 2324; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
OY 1 GTGATCTGTGATCTTTGGCAGAGAGAGGCGGAGCAGAGACAGACATTAATCTGATA 60
Db 1 GTGATCTGTGATCTTTGGCAGAGAGAGGCGGAGCAGAGACAGACATTAATCTGATA 60
OY 61 GTGAGAGCTGGGTTTGGAGATTAAGAGGGAACAGAGAAAGTTAAGTTCTGTGTTTCA 120
Db 61 GTGAGAGCTGGGTTTGGAGATTAAGAGGGAACAGAGAAAGTTAAGTTCTGTGTTTCA 120
OY 121 TGGCAACATTGACAAAAGTTTACACTTCTGTGACTTAACAGTAATCTGGGGTGAATTAC 180
Db 121 TGGCAACATTGACAAAAGTTTACACTTCTGTGACTTAACAGTAATCTGGGGTGAATTAC 180
OY 181 AACAAATTTACACATTAACATATTTACTGACTTTATACAGCAATCTTAAGCTGAAAC 240
Db 181 AACAAATTTACACATTAACATATTTACTGACTTTATACAGCAATCTTAAGCTGAAAC 240
OY 241 ACAGAACCTGCTTATCTTTTGGCAGACTGTCTAGTAGAGATGTCTGTGCTGAGTTA 300
Db 241 ACAGAACCTGCTTATCTTTTGGCAGACTGTCTAGTAGAGATGTCTGTGCTGAGTTA 300
OY 301 AAGAAAGCATTAAGAGCATTAAGTTGTGCACTGTGCAACCCGTGACTTTTTCACACA 360
Db 301 AAGAAAGCATTAAGAGCATTAAGTTGTGCACTGTGCAACCCGTGACTTTTTCACACA 360
OY 361 GTACTAAACCTTAGCTCTTACAGTACAGGAGCATGACGACACAGAAAGAGAAAGCT 420
Db 361 GTACTAAACCTTAGCTCTTACAGTACAGGAGCATGACGACACAGAAAGAGAAAGCT 420
OY 421 CCTTTTACGTGTATATGCTTCTGCTGGGCTTCAAAATCTGTATCTTGAAGATCTGCA 480
Db 421 CCTTTTACGTGTATATGCTTCTGCTGGGCTTCAAAATCTGTATCTTGAAGATCTGCA 480
OY 481 TTACCTGCTTGTGCCCAAAAGTCAATCATCTACCAATGATGTTATTTGAAGTTAA 540
Db 481 TTACCTGCTTGTGCCCAAAAGTCAATCATCTACCAATGATGTTATTTGAAGTTAA 540
OY 541 TCATGTATTAAGAAAGTAGCTACCTCTGAGCCCTGATTAAGAACTTCCCACTGAATACC 600
Db 541 TCATGTATTAAGAAAGTAGCTACCTCTGAGCCCTGATTAAGAACTTCCCACTGAATACC 600
OY 601 GTCTGCTTAAAGTAGAGCAGGTTTCCATTATGATGATGATGAGGAGGAGGAGGAAATA 660
Db 601 GTCTGCTTAAAGTAGAGCAGGTTTCCATTATGATGATGATGAGGAGGAGGAGGAAATA 660
OY 661 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 720
Db 661 TATATATATATATATATATATATATATATATATATATATATATATATATATATATAT 720
OY 721 ACTATCTTTGAAAAGGTTACATATTAACATTTTTCACAGCCAAAGGGGAGAGAA 780
Db 721 ACTATCTTTGAAAAGGTTACATATTAACATTTTTCACAGCCAAAGGGGAGAGAA 780
OY 781 AGATCCAAAAGTCTGTGATCTGCTTAAACATCAATAAAGTATCCACCTTCTGTA 840
Db 781 AGATCCAAAAGTCTGTGATCTGCTTAAACATCAATAAAGTATCCACCTTCTGTA 840
OY 841 GCTTTTATGAAAGGCTACAAAAGTATGCTTTTATGATTAACATGTCAGCGCAACTAC 900
Db 841 GCTTTTATGAAAGGCTACAAAAGTATGCTTTTATGATTAACATGTCAGCGCAACTAC 900
OY 901 TTTTATTTACTACAGAAAAGAGGCTCTTATTAATAAAGGCTTCAAGAACAGTCCAA 960
Db 901 TTTTATTTACTACAGAAAAGAGGCTCTTATTAATAAAGGCTTCAAGAACAGTCCAA 960
OY 961 CAGACTCTGAGGAAATGAAGCAAGTAAATCTGAAAAGGCTTAATAAAGTATGAA 1020
Db 961 CAGACTCTGAGGAAATGAAGCAAGTAAATCTGAAAAGGCTTAATAAAGTATGAA 1020
OY 1021 AATATCTTGTGGAGTTGTTCTTCACTATGCAATAAAGTATATATCATCTATG 1080
Db 1021 AATATCTTGTGGAGTTGTTCTTCACTATGCAATAAAGTATATATCATCTATG 1080
```

```
Db 1021 AATATCTTGTGGAGTTGTTCTTCACTATGCAATAAAGTATATATCATCTATG 1080
OY 1081 ATGGGAAAAACAGGAGCCCTTAATTTCTGAACACCTGTGTGAGAGAGAGGAGG 1140
Db 1081 ATGGGAAAAACAGGAGCCCTTAATTTCTGAACACCTGTGTGAGAGAGAGGAGG 1140
OY 1141 GGCTGCTGCACTCAGAGCGGAGGCTGAGAGGCGGCTCCCTTGCAGAAAGACTGCA 1200
Db 1141 GGCTGCTGCACTCAGAGCGGAGGCTGAGAGGCGGCTCCCTTGCAGAAAGACTGCA 1200
OY 1201 GTGACAGATGGGACACTGAGCTGCCCGGACCTTGAGCCGAGCTGCTAACCTTG 1260
Db 1201 GTGACAGATGGGACACTGAGCTGCCCGGACCTTGAGCCGAGCTGCTAACCTTG 1260
OY 1261 GCCAGGTCCTGCAAGAAATTAAGACCTCCGATTAACCTTAACCACTTCTCACTGCTC 1320
Db 1261 GCCAGGTCCTGCAAGAAATTAAGACCTCCGATTAACCTTAACCACTTCTCACTGCTC 1320
OY 1321 TAAATGTGTGATCCCGGCGCCAGAGGCTTGTGAGCAGAGTGTGCGCTTCCAGGAC 1380
Db 1321 TAAATGTGTGATCCCGGCGCCAGAGGCTTGTGAGCAGAGTGTGCGCTTCCAGGAC 1380
OY 1381 TCCAGCGACCTTAAACCTGAACCGGCGCAGCTCCGACCGGAGGAGAGAAAGAGGC 1440
Db 1381 TCCAGCGACCTTAAACCTGAACCGGCGCAGCTCCGACCGGAGGAGAGAAAGAGGC 1440
OY 1441 ACCGGAGCCCTTCCCGGCGAGCAACCACTTCAACCAAGTTCCGTCAATCCCACT 1500
Db 1441 ACCGGAGCCCTTCCCGGCGAGCAACCACTTCAACCAAGTTCCGTCAATCCCACT 1500
OY 1501 CCTTTCCCGCTCGCAGCGGAGCCAGCTGAGGAGCATGCGCACTGCGAGCGAGGTT 1560
Db 1501 CCTTTCCCGCTCGCAGCGGAGCCAGCTGAGGAGCATGCGCACTGCGAGCGAGGTT 1560
OY 1561 GCCGCGCACAGAGGTAAGTGAATCTGCAACTGTGCGCCCAACCAATCAAGAGA 1620
Db 1561 GCCGCGCACAGAGGTAAGTGAATCTGCAACTGTGCGCCCAACCAATCAAGAGA 1620
OY 1621 CGTGTATTTGCGCGGAGTGGAACTATGCAACGGGAGCAATCAGAAAGCGGTTGT 1680
Db 1621 CGTGTATTTGCGCGGAGTGGAACTATGCAACGGGAGCAATCAGAAAGCGGTTGT 1680
OY 1681 TGCGCGGAGCCCTGCGCCCGGAGGAGGATGTGCGATGAGGATGATGGGCTGT 1740
Db 1681 TGCGCGGAGCCCTGCGCCCGGAGGAGGATGTGCGATGAGGATGATGGGCTGT 1740
OY 1741 GAGCATCCCTGAGCCATGATCCGAGAGGCGCGGGTTCCCTTGCTTGC CGCGGAG 1800
Db 1741 GAGCATCCCTGAGCCATGATCCGAGAGGCGCGGGTTCCCTTGCTTGC CGCGGAG 1800
OY 1801 CGGCGCACAGCCCCGCACTGCTTACCCGAGCCCGGAGCGGCGGCGCCATGCGGC 1860
Db 1801 CGGCGCACAGCCCCGCACTGCTTACCCGAGCCCGGAGCGGCGGCGCCATGCGGC 1860
OY 1861 TGGGAGCGAGGCTTGGAGCGGAGTGTGCGGAGCGGCGGCGCGGAGCGGAGTTGG 1920
Db 1861 TGGGAGCGAGGCTTGGAGCGGAGTGTGCGGAGCGGCGGCGCGGAGCGGAGTTGG 1920
OY 1921 CGGCTGTGCGCGGACTGAAGGCTCGGAGGCGGAGGCGGAGGCGGAGGCGGAGTGT 1980
Db 1921 CGGCTGTGCGCGGACTGAAGGCTCGGAGGCGGAGGCGGAGGCGGAGGCGGAGTGT 1980
OY 1981 GCCGCGGAGCTGAAGAAATGTGCTGTGCTCCGAGTCTCCACAGTAATCCATGCGCTAAC 2040
Db 1981 GCCGCGGAGCTGAAGAAATGTGCTGTGCTCCGAGTCTCCACAGTAATCCATGCGCTAAC 2040
OY 2041 TGCTGCCAGAGTGTGCTCTGAGGCAAGGAGCCCACTTACGTGTGAGCTGAGGAGGCT 2100
Db 2041 TGCTGCCAGAGTGTGCTCTGAGGCAAGGAGCCCACTTACGTGTGAGCTGAGGAGGCT 2100
OY 2101 GCGGCTGTCTCTCGCTTCTGCTGCGCGGCTTCTACCAAGCGCTGAGCAGCGCTGT 2160
Db 2101 GCGGCTGTCTCTCGCTTCTGCTGCGCGGCTTCTACCAAGCGCTGAGCAGCGCTGT 2160
```

QY	2161	ACTGCCGCTCTACCAAGAGATAGTGTGCTCTCTCTCTCTCGAAGATTACACCGGGGATCCAGGTGA	22220
Db	2161	ACTGCCGCTCTACCAAGAGATAGTGTGCTCTCTCTCTCTCGAAGATTACACCGGGGATCCAGGTGA	22220
QY	2221	GCGGCTCTCCGCTCCCGGGTCTCCGGCTCTCCAGCCGAGCTCCCGGGGGGCGGGACCTTCTCC	2280
Db	2221	GCGGCTCTCCGCTCCCGGGTCTCCGGCGTCCAGCCGAGCTCCCGGGGGGCGGGACCTTCTCC	2280
QY	2281	GCTTCCCCCACAAGCTGGCGGAGGGGTACCCCGGCGGGCCCGGGGGAGC	2324
Db	2281	GCTTCCCCCACAAGCTGGCGGAGGGGTACCCCGGCGGGCCCGGGGGAGC	2324
RESULT 3			
BD196565			
LOCUS	BD196565	2000 bp	DNA linear PAT 17-JUN-2003
DEFINITION	Prostatic cancer gene.		
ACCESSION	BD196565		
VERSION	BD196565.1 GI:33006335		
KEYWORDS	JP 2002516657-A/154.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eumayr, J.; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Cohen, D., Blumenfeld, M., Chumakov, I. and Bougueler, L.		
TITLE	Prostatic cancer gene		
JOURNAL	Patent: JP 2002516657-A 154 11-JUN-2002;		
COMMENT	GENSET		
	OS Homo sapiens (human)		
	PN JP 2002516657-A/154		
	PD 11-JUN-2002		
	PF 22-DEC-1998 JP 2000525562		
	PR 22-DEC-1997 US 08/996306, 09-SEP-1998 US 60/099658 PI		
	DANIEL COHEN, MARTA BLUMENFELD, ILYA CHUMAKOV, LYDIE BOUGUELERET PC		
	CI12N5/09, CI12N5/09, A01K67/027, C07K14/47, C07K16/18, CI2N1/15, PC		
	CI2N1/19, PC		
	PC CI2N5/21, CI2N5/10, CI2N5/10, CI2P21/08, CI2Q1/68, G01N33/50 PC		
	, CI2N15/00, CI2N5/00,		
	PC CI2N5/00, CI2N15/00		
	CC Prostatic cancer gene		
	FT key		
	Location/Qualifiers		
	FT source		
	1..2000		
	/organism='Homo sapiens (human)'		
	Location/Qualifiers		
	1..2000		
	/organism='Homo sapiens'		
	/mol_type='genomic DNA'		
	/db_xref='taxon:9606'		
FEATURES			
source			
ORIGIN			
	Query Match	86.1%; Score 2000; DB 6; Length 2000;	
	Best Local Similarity	100.0%; Pred. No. 0;	
	Matches 2000; Conservative	0; Mismatches 0; Indels 0; Gaps 0;	
QY	1	GTGATCTGTGATCTGTTGGCAGGAAGAGAGCGGGAGCAGGACAGACATTAATCTGATA	60
Db	1	GTGATCTGTGATCTGTTGGCAGGAAGAGAGCGGGAGCAGGACAGACATTAATCTGATA	60
QY	61	GTCCAGAGAGCTGGGTTTGGAGATTAAGAGGGGAACAAGAAATTAAGTTCTGTGTTTCA	120
Db	61	GTCCAGAGAGCTGGGTTTGGAGATTAAGAGGGGAACAAGAAATTAAGTTCTGTGTTTCA	120
QY	121	TGGCAAACTTGCACAAAAGTTTACAATCTCGTGACTTAACAGTAATCTGGGGTGAATCA	180
Db	121	TGGCAAACTTGCACAAAAGTTTACAATCTCGTGACTTAACAGTAATCTGGGGTGAATCA	180
QY	181	AACAAAATTTACATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	240
Db	181	AACAAAATTTACATTAACATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	240
QY	241	AACAGAACCTGCTTATCTTTTGACACACTGTTCTAGTGAAGATGTCTGCTCAGTTA	300

[illegible]

```

Db 1321 TAATTGTGTGATCCCGGCGCCAGAGGGCTTGTGTGACAGAGTGTCGGCTTCCAGGAGC 1380
Oy 1381 TCAGCGACCCCTTAACCTGACCGCGCGCATCGCGCCGAGGAGGAGCAACAGAGGC 1440
Db 1381 TCACGACCCCTTAACCTGACCGCGCGCATCGCGCCGAGGAGGAGCAACAGAGGC 1440
Oy 1441 ACCCGGACCCCTTCCCGGCGAGCAGCACCATTACCCAGTTCCGTCAGTCGCGCACACT 1500
Db 1441 ACCCGGACCCCTTCCCGGCGAGCAGCACCATTACCCAGTTCCGTCAGTCGCGCACACT 1500
Oy 1501 CCGTTCCCGGCGTCCGCGAGCCGCGAGCTGGGAGAGCATGCGCAGTGGCCGAGCGGTT 1560
Db 1501 CCGTTCCCGGCGTCCGCGAGCCGCGAGCTGGGAGAGCATGCGCAGTGGCCGAGCGGTT 1560
Oy 1561 GCGCGCGCGACAGAGGTAGTGTCTGCAACTGTGCGCCCAACCAACCAATCAAGAGA 1620
Db 1561 GCGCGCGCGACAGAGGTAGTGTCTGCAACTGTGCGCCCAACCAACCAATCAAGAGA 1620
Oy 1621 CGTGTATTGCGCGCGAGGTGAGACTATGCAACCGGCGCGCAATCAGAAAGCGGTTGT 1680
Db 1621 CGTGTATTGCGCGCGAGGTGAGACTATGCAACCGGCGCGCAATCAGAAAGCGGTTGT 1680
Oy 1681 TGCCGCGGAGCCCTTCCCGGCGAGGAGGATGTGGCGATGGGTGATGGGGTGT 1740
Db 1681 TGCCGCGGAGCCCTTCCCGGCGAGGAGGATGTGGCGATGGGTGATGGGGTGT 1740
Oy 1741 GAGCATCCCTGAGCATGATCCGGGAGGCGCGGGGTTCCCTTGTGCGCGCGGAG 1800
Db 1741 GAGCATCCCTGAGCATGATCCGGGAGGCGCGGGGTTCCCTTGTGCGCGCGGAG 1800
Oy 1801 CGGCGCGACGAGCCCGGCGACTGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCG 1860
Db 1801 CGGCGCGACGAGCCCGGCGACTGCGCTTACCGCGCGCGCGCGCGCGCGCGCGCG 1860
Oy 1861 TGGGGGCGGAGGCTGGGAGCGGGGTGGCGCGCGCGCGCGCGCGCGCGCGGTATTGG 1920
Db 1861 TGGGGGCGGAGGCTGGGAGCGGGGTGGCGCGCGCGCGCGCGCGCGCGCGGTATTGG 1920
Oy 1921 CCGCGCTGTGGCGCGGACTGAGGCGCGGGAGGCGGGAGGCGGAGCGGAGGCGGAGTCCCT 1980
Db 1921 CCGCGCTGTGGCGCGGACTGAGGCGCGGGAGGCGGGAGGCGGAGCGGAGGCGGAGTCCCT 1980
Oy 1981 GCGCGCGGAGCTGAGGAGGATG 2000
Db 1981 GCGCGCGGAGCTGAGGAGGATG 2000

RESULT 4
AC100813/c 165799 bp DNA linear HTG 27-MAR-2003
LOCUS Homo sapiens chromosome 8 clone CTD-2541M15 map 8, 4 unordered
DEFINITION Homo sapiens chromosome 8 clone CTD-2541M15 map 8, 4 unordered
pieces.
AC100813
VERSION AC100813.2 GI:22381986
KEYWORDS HTG; HTGS PHASE1; HTGS_FULLTOP; HTGS_CANCELLED.
SOURCE Homo sapiens (human)
ORANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 165799)
Britten,B., Nusbau,C. and Lander,B.
2 (bases 1 to 165799)
unpublished
Britten,B., Linton,L., Nusbau,C., Lander,B., Ali,A., Allen,N.,
Anderson,S., Barna,N., Bastien,V., Boguslavsky,L., Boukhaltier,B.,
Brown,A., Camarata,J., Campopiano,A., Chang,J., Chazaro,B.,
Choepel,Y., Colangelo,M., Collins,S., Collymore,A., Cook,A.,
Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S., Faro,S.,
Ferreira,P., Fitzhugh,W., Gage,D., Galagan,J., Gardyna,S.,
Ginde,S., Gord,S., Goyette,M., Graham,L., Grand-Pierre,N.,
Hagos,B., Heaford,A., Horton,L., Hulme,W., Iliev,I., Johnson,R.,
Jones,C., Kamat,A., Karatas,A., Kelle,C., LaRocque,K.,

```

```

Lamarez,R., Landers,T., Lehoczy,J., Levine,R., Liu,G.,
Maclean,C., Macdonald,P., Major,J., Margulis,N., Mathews,C.,
McCarthy,M., Mcwan,P., McKernan,K., McPheters,R., Meldrim,J.,
Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J., Nguyen,C.,
Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neill,D.,
Oliver,J., Peterson,K., Phunkhang,P., Pierre,N., Pollara,V.,
Raymond,C., Retta,R., Rieback,M., Riley,R., Rise,C., Rogov,P.,
Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schuppach,R.,
Seaman,S., Severy,P., Spencer,B., Stange-Thomann,N., Stojanovic,N.,
Strauss,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Topham,K., Travers,M., Travis,N., Trigglio,J., Vasiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G.,
Zainoun,J., Zembek,L., Zimmer,L., and Zody,M.
Direct Submission
Submitted (22-NOV-2001) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 165799)
Britten,B., Nusbau,C., Lander,B., Ali,A., Allen,N., Anderson,S.,
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhaltier,B.,
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
Faro,S., Ferreira,P., Fitzerald,M., Gage,D., Galagan,J.,
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,
Karatas,A., Kelle,C., Landers,T., Levine,R., Lindblad-Toh,K.,
Liu,G., Maclean,C., Macdonald,P., Major,J., Mathews,C.,
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,
Roman,J., Roy,A., Schauer,S., Schuppach,R., Seaman,S., Severy,P.,
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vasiliev,H.,
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,
Zembek,L., Zimmer,L., and Zody,M.
Direct Submission
Submitted (27-MAR-2003) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Aug 21, 2002 this sequence version replaced gi:117048183.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
http://ftp.genome.washington.edu/RM/RepeatMasker.html
----- Genome Center
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WtBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence_submissions@genome.wi.mit.edu
----- Project Information
Center project name: L21488
Center clone name: 2541_M_15
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 4 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 62727: contig of 62727 bp in length
* 2 62827: gap of 100 bp
* 3 73470: contig of 10643 bp in length
* 4 73471: gap of 100 bp
* 5 73571: contig of 7945 bp in length
* 6 81516: gap of 100 bp
* 7 81516: contig of 84184 bp in length.
*
* Location/Qualifiers
* 1. 165799
* /organism="Homo sapiens"
* /mol_type="genomic DNA"
* /db_xref="taxon:9606"
* /chromosome="8"
* /map="8"

```


TITLE Direct Submission
JOURNAL Submitted (23-AUG-2001) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
REFERENCE 4 (bases 1 to 168136)
AUTHORS Lagemann,D. and Platzer,M.
JOURNAL Direct Submission
Submitted (29-MAR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
REFERENCE 5 (bases 1 to 168136)
AUTHORS Lagemann,D. and Platzer,M.
JOURNAL Direct Submission
Submitted (03-APR-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena, Thuringia 07745, Germany
REFERENCE 6 (bases 1 to 168136)
AUTHORS Lagemann,D. and Platzer,M.
JOURNAL Direct Submission
Submitted (09-JUL-2002) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
REFERENCE 7 (bases 1 to 168136)
AUTHORS Lagemann,D. and Platzer,M.
JOURNAL Direct Submission
Submitted (08-JUL-2003) Genome Analysis, Institute of Molecular
Biotechnology, Beutenbergstr. 11, Jena 07745, Germany
On Jul 8, 2003 this sequence version replaced gi:21717104.
COMMENT Genome Center
Center: Institute of Molecular Biotechnology
Center code: IMB
Web site: <http://genome.imb-jena.de/>
Contact: gsfj-submit@genome.imb-jena.de
Project Information
Center project name: H351
Center clone name: CTD-2541M15
Summary Statistics
Sequencing vector: pUC18; 100% of reads
Chemistry: Dye-terminator Big Dye; 100% of reads
Assembly program: Phrap; version 0.990329
Consensus quality: 167707 bases at least Q40
Consensus quality: 168123 bases at least Q30
Consensus quality: 168136 bases at least Q20
Quality coverage: 6.18x

This sequence was finished as follows unless otherwise noted: all
regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by at least
one plasmid subclone or more than one pUC18 subclone; and the
assembly was confirmed by restriction digest.

Neighboring sequence information:
This clone is overlapped by GSI-24F4, GSI-24F4.

Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.

FEATURES
SOURCE
1..168136
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="p22-p21"
/clone="CTD-2541M15"
165266..168136
/organism="Homo sapiens"
/mol_type="genomic DNA"

/db_xref="taxon:9606"
/clone="GSI-24F4"
/note="overlapping clone, overlapping clone"
6418
misc_feature
/note="low quality region"
7442..7645
misc_feature
/note="single stranded/single chemistry region"
28243..28507
misc_feature
/note="single stranded/single chemistry region"
28499
misc_feature
/note="low quality region"
28501
misc_feature
/note="low quality region"
34532..34843
misc_feature
/note="single stranded/single chemistry region"
35987..36115
misc_feature
/note="single stranded/single chemistry region"
37530
misc_feature
/note="low quality region"
38752..38820
misc_feature
/note="single stranded/single chemistry region"
38995..39017
misc_feature
/note="single stranded/single chemistry region"
44702..44825
misc_feature
/note="single stranded/single chemistry region"
45312..45406
misc_feature
/note="single stranded/single chemistry region"
47601..47995
misc_feature
/note="single stranded/single chemistry region"
59027..59136
misc_feature
/note="single stranded/single chemistry region"
61319
misc_feature
/note="low quality region"
62590..62726
misc_feature
/note="single stranded/single chemistry region"
66851..66899
misc_feature
/note="single stranded/single chemistry region"
69911..70188
misc_feature
/note="single stranded/single chemistry region"
75977..76021
misc_feature
/note="single stranded/single chemistry region"
78279..78386
misc_feature
/note="single stranded/single chemistry region"
78675
misc_feature
/note="low quality region"
81011..81279
misc_feature
/note="single stranded/single chemistry region"
81070..81071
misc_feature
/note="low quality region"
81743..82001
misc_feature
/note="single stranded/single chemistry region"
81872
misc_feature
/note="low quality region"
84434..85028
misc_feature
/note="pcr product sequence only"
84932..85028
misc_feature
/note="single stranded/single chemistry region"
86073..86097
misc_feature
/note="single stranded/single chemistry region"
86609
misc_feature
/note="low quality region"
89636..89783
misc_feature
/note="single stranded/single chemistry region"
90439..90940
misc_feature
/note="single stranded/single chemistry region"
95593..95693
misc_feature
/note="single stranded/single chemistry region"
98827..98848
misc_feature
/note="single stranded/single chemistry region"
98860
misc_feature
/note="low quality region"
99495..99715
misc_feature
/note="single stranded/single chemistry region"

QY	914	GAAGAAAAAGAGGCTCCTTTATTTAAAAAAAATCAGAAA	CAAGTCCAA	CGA	CTCGAGGA	973
Db	73173	GAAGAAAAAGAGGCTCCTTTATTTAAAAAAAATCAGAAA	CAAGTCCAA	CGA	CTCGAGGA	73232
QY	974	AATGAAGCAAGATGATATTTCTGAAAAAGTCTAATTAACA	GTATGAAATAT	CTCTGTGGG		1033
Db	73233	AATGAAGCAAGATGATATTTCTGAAAAAGTCTAATTAACA	GTATGAAATAT	CTCTGTGGG		73292
QY	1034	ATTGTTCTTCAGCTATGCAATAAACATGTAATTAATCAT	TCATTA	CTGTGATGGG	AAAAACA	1093
Db	73293	ATTGTTCTTCAGCTATGCAATAAACATGTAATTAATCAT	TCATTA	CTGTGATGGG	AAAAACA	73352
QY	1094	CGAGCCCTAATTCTGAAAACACCTGTGATCGAGAGA	CGGCGAGAG	GCGGCTGCGAC		1153
Db	73353	CGAGCCCTAATTCTGAAAACACCTGTGATCGAGAGA	CGGCGAGAG	GCGGCTGCGAC		73412
QY	1154	TCGAGAGCGAGGCTGAGAGAGCGGCGCTCCCTTGCAAA	AGAC	CTGGAGTGA	CAACATGGG	1213
Db	73413	TCGAGAGCGAGGCTGAGAGAGCGGCGCTCCCTTGCAAA	AGAC	CTGGAGTGA	CAACATGGG	73472
QY	1214	GACACTCGACCTGCGCCGCGACCTGGGCGAGCTGCTCA	CAACTGGG	CCGAGTGCCTG		1273
Db	73473	GACACTCGACCTGCGCCGCGACCTGGGCGAGCTGCTCA	CAACTGGG	CCGAGTGCCTG		73532
QY	1274	CAAGAATTAAGACCTTCGCAATAACGTAAACACCACTT	TTCTAC	TGCTTAATGTGTGAT		1333
Db	73533	CAAGAATTAAGACCTTCGCAATAACGTAAACACCACTT	TTCTAC	TGCTTAATGTGTGAT		73592
QY	1334	CCCGGCGCCCGAGGSGGCTTGTGAGAGAGGTGCGCTT	CCAGGCA	CTCCAGGACCTT		1393
Db	73593	CCCGGCGCCCGAGGSGGCTTGTGAGAGAGGTGCGCTT	CCAGGCA	CTCCAGGACCTT		73652
QY	1394	AAACCTTAACGCGCGACAGTCCCGGCGGAGGAGACAA	AGAGGCA	CCCGGACCTTCC		1453
Db	73653	AAACCTTAACGCGCGACAGTCCCGGCGGAGGAGACAA	AGAGGCA	CCCGGACCTTCC		73712
QY	1454	TCCGGCCAGACCCCACTTCAACCCAGTTCCGTACAGT	GGCACAC	CTCCCTTCCGCGTCC		1513
Db	73713	TCCGGCCAGACCCCACTTCAACCCAGTTCCGTACAGT	GGCACAC	CTCCCTTCCGCGTCC		73772
QY	1514	CGCAGCCGGCCAGACTGGGAGACATGGCAGTGGCCGG	AGCCG	GTGGCCCGGCGCACAG		1573
Db	73773	CGCAGCCGGCCAGACTGGGAGACATGGCAGTGGCCGG	AGCCG	GTGGCCCGGCGCACAG		73832
QY	1574	CAGGTAGCTGTACATGCAACTGTGGGCCCAAAACCA	ACCAATCA	AGAGAGCGTATTATGGCG		1633
Db	73833	CAGGTAGCTGTACATGCAACTGTGGGCCCAAAACCA	ACCAATCA	AGAGAGCGTATTATGGCG		73892
QY	1634	CCGAGGTGGAATCTATGGCAACGGGCGACCAATCA	GAAAGGCG	CTGTGTGGCCGGAGCC		1693
Db	73893	CCGAGGTGGAATCTATGGCAACGGGCGACCAATCA	GAAAGGCG	CTGTGTGGCCGGAGCC		73952
QY	1694	CCTGCCCCCGGAGGGGGATGTGGCGATGGGTGAGAG	GTCAATCC	CTGTAGG		1753
Db	73953	CCTGCCCCCGGAGGGGGATGTGGCGATGGGTGAGAG	GTCAATCC	CTGTAGG		74012
QY	1754	CCATTCGATCCCGGAGGGGCGCGGCTTCCCTTGCTTT	CCGCGCGG	AGCGGCGACG	CAACG	1813
Db	74013	CCATTCGATCCCGGAGGGGCGCGGCTTCCCTTGCTTT	CCGCGCGG	AGCGGCGACG	CAACG	74072
QY	1814	CCCGCACTCGCTTACC	CGGCCCCCGGCGCGGCGCGGCG	CCATGCGG	CTGGGGGCGGAGG	1873
Db	74073	CCCGCACTCGCTTACC	CGGCCCCCGGCGCGGCGCGGCG	CCATGCGG	CTGGGGGCGGAGG	74133
QY	1874	TGGGAGCGGAGTGGCGGCGCGGCGCGGCGCGGCG	CGGAGTGG	CAATGG	CCGCTGGGCGGAGG	1933
Db	74133	TGGGAGCGGAGTGGCGGCGCGGCGCGGCGCGGCG	CGGAGTGG	CAATGG	CCGCTGGGCGGAGG	74192
QY	1934	GCAGCTTGAAGCCCGGAGGCGGAGCGGAGAGCGCA	GGCGGAG	CTGCTG	CCGCGGAGCTGA	1993
Db	74193	GCAGCTTGAAGCCCGGAGGCGGAGCGGAGAGCGCA	GGCGGAG	CTGCTG	CCGCGGAGCTGA	74252
QY	1994	GAAGATGCTGCTCTCCCTGTGTGCTCCACAGCTAT	CCATGCGG	CTACTG	CTGCTGCGACG	2053

Db	74253	AAAGATGCTGCTGTCCCTGAGTCTCCACACGTACTCAATGCGCTTACCTGCTGCAGAGGT	74312
Qy	2054	CGTGTCTCTGGGACACGGCGGCCACCTAAGTGTGGCTCGGGGGGGCTGGCGGCTGCTCTC	2113
Db	74313	CGTGTCTCTGGGACACGGCGGCCACCTAAGTGTGGCTCGGGGGGGCTGGCGGCTGCTCTC	74372
Qy	2114	CGGCTTCTGCGCCGCGCCGCTTCTAACCAACGCGTGGACGACCGGCTGTACTGGCTTACCA	2173
Db	74373	CGGCTTCTGCGCCGCGCCGCTTCTAACCAACGCGTGGACGACCGGCTGTACTGGCTTACCA	74432
Qy	2174	GAGCATGTGTCTCTTCTTCTTGAGATTTACACCGGGGTCGAGTGAAGCCGCTCCCGT	2233
Db	74433	GAGCATGTGTCTCTTCTTCTTGAGATTTACACCGGGGTCGAGTGAAGCCGCTCCCGT	74492
Qy	2234	CCCGGGTCTTGGGGGTCCACCCGAGCTCCGGGGGGCGGACCTCTCCGCTCCCCACAGC	2293
Db	74493	CCCGGGTCTTGGGGGTCCACCCGAGCTCCGGGGGGCGGACCTCTCCGCTCCCCACAGC	74552
Qy	2294	TGGGAGGGGTCAACCGGCGGACCGGCGGAC	2324
Db	74553	TGGGAGGGGTCAACCGGCGGACCGGCGGAC	74583

RESULT 7	209718 bp	DNA	linear	HTG 06-AUG-2002
AC091099	Homo sapiens chromosome 8 clone RP11-782N1	map 8,	WORKING DRAFT	
LOCUS	SEQUENCE, 10 unordered pieces.			
DEFINITION				
ACCESSION	AC091099.3	GI:22122960		
VERSION	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_FULLTOP.			
KEYWORDS	Homo sapiens (human)			
SOURCE	Homo sapiens			
ORGANISM	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.			
REFERENCE	1 (bases 1 to 209718)			
AUTHORS	Birren, B., Nusbaum, C. and Lander, E.			
TITLE	Homo sapiens chromosome 8, clone RP11-782N1			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 209718)			
AUTHORS	Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,			

TITLE Direct Submission
JOURNAL Submitted (29-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE 3 (bases 1 to 209718)
AUTHORS Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S.,

Maclean, C., Macdonald, P., Margulis, N., Matthews, C., McCarthy, M., McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Meneus, L., Mihova, T., Mleczka, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Plunhachang, P., Pierre, N., Pollara, V., Raymond, C., Retta, R., Riebeck, M., Riley, R., Rise, C., Rogov, P., Roman, J., Rossetti, M., Roy, A., Santos, R., Schauer, S., Schuppach, R., Seaman, S., Severy, P., Sounguez, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Travers, M., Travis, N., Triggilio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zaimoun, J., Zembek, L., Zimmer, A. and Zody, M.

Direct Submission
Submitted (29-MAR-2001) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 209718)

Birren, B., Nusbaum, C., Lander, E., Ali, A., Allen, N., Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguski, L., Boukanger, B., Camarata, J., Chang, J., Chazaro, B., Chopel, Y., Collamore, A., Cook, A., Cooke, P., Detrellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, K., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kanati, A.,

Db 127894 CAAGATTAGACCTCCGATTAACGTTAAGACCACTTTCTCACTGCTTAATTGTGTGAT 127953
 QY 1334 CCGGGGCGCCAGGGGCTTTGAGCAGAGTGGGGTTCCAGGAGCTCCAGGACCTT 1393
 Db 127954 CCGGGGCGCCAGGGGCTTTGAGCAGAGTGGGGTTCCAGGAGCTCCAGGACCTT 128013
 QY 1394 AAACCTGACCGGCGCAGCTCCGGCCCGAGGAGAGAACAGAGAGCACCAGGACCTTC 1453
 Db 128014 AAACCTGACCGGCGCAGCTCCGGCCCGAGGAGAGAACAGAGAGCACCAGGACCTTC 128073
 QY 1454 TCCGGCCAGCAGCACCACCTTCACTCCAGTTCCGTCACTCCGACCACTCCCTTCCGGCTC 1513
 Db 128074 TCCGGCCAGCAGCACCACCTTCACTCCAGTTCCGTCACTCCGACCACTCCCTTCCGGCTC 128133
 QY 1514 CGCAGCCGGCCGACGCTGGGGAGACATAGCGAGTGGCGGAGCGGGTTGGCCGCGCAGCAG 1573
 Db 128134 CGCAGCCGGCCGACGCTGGGGAGACATAGCGAGTGGCGGAGCGGGTTGGCCGCGCAGCAG 128193
 QY 1574 CAGGTAGCTGTACTGCAACTGTGCGCCCAACCAACCAATCAAGAGACGTGTTATTGCGC 1633
 Db 128194 CAGGTAGCTGTACTGCAACTGTGCGCCCAACCAACCAATCAAGAGACGTGTTATTGCGC 128253
 QY 1634 CCGAGGTGGAATCTATGAGCAGCGGCGGACCAATCAAGAGCGCGTTGTTCCCGGAGCC 1693
 Db 128254 CCGAGGTGGAATCTATGAGCAGCGGCGGACCAATCAAGAGCGCGTTGTTCCCGGAGCC 128313
 QY 1694 CCGTGGCCCGGAGGGGGAGTGTGGCGCATGGGTGAGGGGTATGGGGTGTGAGCATCCCTGAG 1753
 Db 128314 CCGTGGCCCGGAGGGGGAGTGTGGCGCATGGGTGAGGGGTATGGGGTGTGAGCATCCCTGAG 128373
 QY 1754 CCATCGATCCGGAGAGGGCGCGGGGTTCCCTTCCGCGGAGCGCGCAGCAGCAGC 1813
 Db 128374 CCATCGATCCGGAGAGGGCGCGGGGTTCCCTTCCGCGGAGCGCGCAGCAGCAGC 128433
 QY 1814 CCGGCACTGCTCTATCCCGGCCCCGGGGCGGGCGGCGGCGGCAATGCGGTGGGGGCGGAGGC 1873
 Db 128434 CCGGCACTGCTCTATCCCGGCCCCGGGGCGGGCGGCGGCGGCAATGCGGTGGGGGCGGAGGC 128493
 QY 1874 TGGGAGCGGGGTGGCGGGCGGGCGGGCGGGCGGGCGGGTGTGATGGCGGCTGCTGAGCC 1933
 Db 128494 TGGGAGCGGGGTGGCGGGCGGGCGGGCGGGCGGGCGGGCGGGTGTGATGGCGGCTGCTGAGCC 128553
 QY 1934 GCGACTGAGCGCCCGGAGGGCGGGCGGGGAGCGGAGCGGAGCTCGCTGCGCGGAGCTGA 1993
 Db 128554 GCGACTGAGCGCCCGGAGGGCGGGCGGGGAGCGGAGCGGAGCTCGCTGCGCGGAGCTGA 128613
 QY 1994 GAAAGTGTGCTGTCTCCCTGTGTCTCCACACGTAATCCATGCGCTACCTGTCCAGCGT 2053
 Db 128614 AAGAGTGTGCTGTCTCCCTGTGTCTCCACACGTAATCCATGCGCTACCTGTCCAGCGT 128673
 QY 2054 CGTGTCTCTGGGCAAGCGGCGGCACTACGTGTTGGCGCTGGGGGGTCTGGCGGCTCTC 2113
 Db 128674 CGTGTCTCTGGGCAAGCGGCGGCACTACGTGTTGGCGCTGGGGGGTCTGGCGGCTCTC 128733
 QY 2114 GCGCTTCTCTGCGCGCGGCTTCTACCAAGCGCTGAGCGAGCGGCTGTACTGCGTCTACCA 2173
 Db 128734 GCGCTTCTCTGCGCGCGGCTTCTACCAAGCGCTGAGCGAGCGGCTGTACTGCGTCTACCA 128793
 QY 2174 GAGCATGTGCTCTTCTTCTTGAAGATTAACAACCGGGGTCAAGTGAAGCGGCTCCCGCT 2233
 Db 128794 GAGCATGTGCTCTTCTTCTTGAAGATTAACAACCGGGGTCAAGTGAAGCGGCTCCCGCT 128853
 QY 2234 CCGGGGTCTCGGCGTCAACCGGAGCTCCGGGGGCGGAGCTCTCGCTCCCGCAGCAGC 2293
 Db 128854 CCGGGGTCTCGGCGTCAACCGGAGCTCCGGGGGCGGAGCTCTCGCTCCCGCAGCAGC 128913
 QY 2294 TGGCGAGGCTCAACCGGCGGCGCGGCGGAGC 2324
 Db 128914 TGGCGAGGCTCAACCGGCGGCGCGGCGGAGC 128944

RESULT 8
 AC009632/c

AC009632 185634 bp DNA linear HTG 07-JUN-2000
 Homo sapiens chromosome 8 clone RP11-337K22 map 8, *** SEQUENCING
 IN PROGRESS ***, 48 unordered pieces.
 AC009632
 AC009632.3 GI:8954116
 HTG: HTG_PHASE1
 SOURCE
 Homo sapiens (human)
 ORGANISM
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 185634)
 Birren, B., Linton, L., Nusbaum, C. and Lander, E.
 Homo sapiens chromosome 8, clone RP11-397K22
 Unpublished
 2 (bases 1 to 185834)
 Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, M.,
 Baker, J., Baldwin, J., Barna, N., Beckert, R., Benn, J., Brown, A.,
 Castle, A., Cerny, J., Colangelo, M., Collins, S., Collamore, A.,
 Cooke, P., DeArrellano, K., Depayre, B., Devon, K., Dewar, K.,
 Donelan, L., Doyle, M., Ferreira, P., Fitzhugh, W., Forrest, C.,
 Funke, R., Gage, D., Galagan, J., Gardyna, S., Gilbert, D., Grant, G.,
 Hacos, B., Heaford, A., Horton, L., Howland, J. C., Jones, C., Kann, L.,
 Karatas, A., Lehoczy, J., Lieu, C., Locke, K., MacDonald, P.,
 Marquis, N., McEwan, P., McGurk, A., McKernan, K., McLaughlin, J.,
 Meidrum, J., Molla, M., Morris, W., Morrow, J., Mychaleckyj, J.,
 Naylor, J., Nijhoff, M., O'Connor, T., O'Donnell, P., Pavlin, B.,
 Peterson, K., Pollara, V., Riley, R., Roberts, D., Roy, A., Severy, P.,
 Stange-Thomann, N., Stojanovic, N., Stone, C., Subramanian, A.,
 Teefaye, S., Tornetta-Miller, I., Vassiliev, H., Vo, A., Wagner, A.,
 Wheeler, J., Wu, X., Wyman, D., Ye, W. J. and Zody, M.
 Direct Submission
 Submitted (28-AUG-1999) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 On Jul 7, 2000 this sequence version replaced gi:5932606.
 All repeats were identified using RepeatMasker:
 Smit, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html
 ----- Genome Center
 Center: Whitehead Institute/ MIT Center for Genome Research
 Center code: WIRB
 Web site: http://www-seq.wi.mit.edu
 Contact: sequence_submissions@genome.wi.mit.edu
 ----- Project Information
 Center project name: L2099
 Center clone name: 397_K_22

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 48 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.
 * 1
 * 1362 1461: contig of 1361 bp in length
 * 1462 3202: contig of 1741 bp in length
 * 3203 3302: gap of 100 bp
 * 3303 4619: contig of 1317 bp in length
 * 4620 4719: gap of 100 bp
 * 4720 5821: contig of 1102 bp in length
 * 5822 5921: gap of 100 bp
 * 5922 6419: contig of 498 bp in length
 * 6420 6519: gap of 100 bp
 * 6520 8242: contig of 1723 bp in length
 * 8243 8342: gap of 100 bp
 * 8343 9703: contig of 1361 bp in length
 * 9704 9803: gap of 100 bp
 * 9804 10936: contig of 1133 bp in length
 * 10937 11036: gap of 100 bp
 * 11037 12339: contig of 1303 bp in length
 * 12340 12440 12440: gap of 100 bp in length
 * 12440 13923: contig of 1484 bp in length

Oy		2033	GCGCTACCTGCTGCACCAGCGTCGAGTCTCTTG36GACGGCGGCCACTTACGTGTGGCCTG	2092
Dd		310	GCCTCACTGCTGTCCCAAGCTGTGCTCTTG36GACGGCGGCCACTTACGTGTGGCCTG	369
Oy		2093	GGGGGCTGAGCGGCTCTCTCGCCTTCCTGCGCCCGCTTTCTTACCAAGCGCTGAAGA	2152
Dd		370	GGGGGCTGAGCGGCTCTCTCGCCTTCCTGCGCCCGCTTTCTTACCAAGCGCTGAAGA	429
Oy		2153	CCGCGCTGACCTGCTCTTACCGAGCATGTGCTCTTCTTTCCAGAAATTACCGCGGCT	2212
Dd		430	CCGCGCTGACCTGCTCTTACCGAGCATGTGCTCTTCTTTCCAGAAATTACCGCGGCT	489
Oy		2213	CCAG	2216
Dd		490	CCAG	493
	RESULT 10			
	AF375789			
	LOCUS			
	DEFINITION	AF375789	1509 bp	mRNA linear PRI 21-MAY-2001
	ACCESSION	(LPAAT-e)	mRNA, complete cds.	
	VERSION	AF375789		
	KEYWORDS	AF375789.1	GI:14161584	
	SOURCE	Homo sapiens (human)		
	ORGANISM	Homo sapiens		
	REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
	AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
	TITLE	1 (bases 1 to 1509)		
	JOURNAL	Cloning and expression of LPAAT-epsilon		
	REFERENCE	2 (bases 1 to 1509)		
	AUTHORS	Leung, D.W.		
	JOURNAL	Direct Submission		
	TITLE	Submitted (02-MAY-2001) Mol. Biol., Cell Therapeutics, Inc., 201		
	FEATURES	Location/Qualifiers		
	source	1..1509		
		/organism="Homo sapiens"		
		/mol_type="mRNA"		
		/db_xref="taxon:9606"		
		/chromosome="8"		
		/map="gpter-p23.3"		
		1..1509		
		/gene="LPAAT-e"		
		275..1369		
		/gene="LPAAT-e"		
		/EC_number="2.3.1.51"		
		/note="1-acylglycerol-3-phosphate O-acyltransferase"		
		/codon_start=1		
		/product="lysophosphatidic acid acyltransferase-epsilon"		
		/protein_id="AAK54809.1"		
		/db_xref="gi:14161583"		
		/translation="MLSLVHTYTSMYRLPSVVLGTAPTYIYLAQGWRLLSAFLPRA RFPYALDRLICVVOQMWLFFPENYTVGVQILLYSDLPKNKENIIYLANHOSTVDIMIVYA DIILAIROMALDRIVRYLVKEGLMKPLVGYCYFAOHGGIYVKRSAKNEKEMNKQSYYVY DAGRPVLTIVIPREGRYNPEQTIVLASQAFAORGLAVTHVLTPIRIKATHVAEDCK KNYDAIYDVTVVVYEGKDGGCORRESPTMTFPLCKECKIHIDRIDDKVPPGEQER MRRLHERFEIKDKMLLEFTSPDEPRKRPGKSVNSKSIKTLPMLITLSGLTNG MLMDAGRKLIYNVTWITYGTLLGLMWTIKA"		
	gene			
	CDS			
	ORIGIN			
	Query Match	18.6%	Score 433;	DB 9; Length 1509;
	Best Local Similarity	99.8%;	Pred. No. 4.3e-235;	
	Matches 483; Conservative	0;	Mismatches 1;	Indels 0; Gaps 0;
Oy		1733	TGGGGTGTGACATCCCTGAGCCATCGATCCGGAGAGGCGCGGCTTCCCTTGCTTGGCC	1792
Dd		10	TGGGGTGTGACATCCCTGAGCCATCGATCCGGAGAGGCGCGGCTTCCCTTGCTTGGCC	69
Oy		1793	GCCGGAGCGCGCACGACCCCGCACTCGCTACCCGGCCCCGGGCGGCGCGCGGCC	1852

D	b70	GCCGGAGCGGCGCACCACGCCATCTGCCTTACCOCGCCCCGGCGCGGCGCCC	129
OY	1853	CATCGCGGTGGGAGCGCGAAGCTTGAGCGGTCGGCGCGCGCGCCCGGCG	1912
D	b130	CATCGCGGTGGGAGCGCGAAGCTTGAGCGGTCGGCGCGCGCGCCCGGCG	189
OY	1913	GTAATTGGCGCGCTGCTCGGCGCGGAAGAATACTGTTCCCTGGTCTCACACTCA	1972
D	b190	GTAATTGGCGCGCTGCTCGGCGCGGAAGAATACTGTTCCCTGGTCTCACACTCA	249
OY	1973	AGCTCGTGGCGCGCGAAGCTGAGAAGTAGCTGCTGCTCCCTGGTCTCACACTCA	2032
D	b250	AGCTCGTGGCGCGCGAAGCTGAGAAGTAGCTGCTGCTCCCTGGTCTCACACTCA	309
OY	2033	GCGCTACCTGCTGCCAGCGTCTGCTCCTTGAGCGCGCGCCACAATTGAGCTTG	2092
D	b310	GCGCTACCTGCTGCCAGCGTCTGCTCCTTGAGCGCGCGCCACAATTGAGCTTG	369
OY	2093	GCGGCTGCGCGGCTGCTCTCCGCGCTTCCTGCGCGCGCGCTTCTCATCAAACGCTGAG	2152
D	b370	GCGGCTGCGCGGCTGCTCTCCGCGCTTCCTGCGCGCGCGCTTCTCATCAAACGCTGAG	429
OY	2153	CCGCGCTTACCTGCGGTACACAGACATGATGCTCTTCTTGAGAAATTACACCGAGGT	2212
D	b430	CCGCGCTTACCTGCGGTACACAGACATGATGCTCTTCTTGAGAAATTACACCGAGGT	489
OY	2213	CCAG 2216	
D	b490	CCAG 493	
R	E SULT 11		
L	B C O 2 3 5 5 0	1503 bp mRNA linear PRI 06-OCT-2003	
I	N D E F I N I T I O N	Homo sapiens acid acyltransferase-beta, complete cds.	
A	C C E S S I O N	MGC:19506 IMAGE:412973), complete cds.	
V	E R S I O N	BC023550 GI:33879276	
K	E Y W O R D S	MGC.	
S	T A X O N O M Y	Homo sapiens (human)	
O	R G A N I S M	Homo sapiens	
R	E F E R E N C E	Bukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
A	U T H O R S	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.	
		1 (bases 1 to 1503)	
		Strausberg,R.L., Feingold,E.A., Grouse,L.H., Dege,J.G., Klausner,R.D., Collins,F.S., Wagner,K.H., Shenmen,C.M., Schuler,G.D., Altschuler,S.F., Zeeberg,B., Bluelow,K.H., Scheaffer,C.F., Bhat,N.K., Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F., Diatchenko,M., Soares,M.B., Bonaldi,M.F., Casavant,T.J., Scheerz,T.E., Brownstein,M.J., Ushin,T.B., Toshiyuki,S., Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J., Abramson,R.D., Mulvihy,S.J., Bosak,S.A., McEwan,P.J., McKernan,K.O., Malek,J.A., Gunaratne,P.H., Richards,S., Morley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulik,S.W., Viallalon,D.R., Munzy,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A., Fahney,J., Helton,E., Kettman,M., Madan,A., Rodriguez,S., Sanchez,A., Whitting,M., Madan,A., Young,A.C., Shevchenko,Y., Bouffard,G.G., Blakeley,R.W., Touchman,J.W., Green,E.D., Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M., Buttefield,Y.S., Krzywinski,M.I., Skalske,U., Smalls,D.E., Schneider,A., Schen,J.E., Jones,S.J., and Marra,M.A.	
		Generation and initial analysis of more than 15,00 full-length human and mouse cDNA sequences Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)	
J	O U R N A L	Medline 12388257	
P	U B M E D	12477932	
R	E F E R E N C E	2 (Bases 1 to 1503)	
A	U T H O R S	Strausberg,R.	
T	I T L E	Direct Submission	
J	O U R N A L	Submitted (05-FEB-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer	

REMARK
COMMENT

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
NIH-MGC Project URL: <http://mgc.nci.nih.gov>
On Aug 19, 2003 this sequence version replaced gi:23272119.
Contact: MGC help desk
Email: cgabbs-remail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: National Institutes of Health Intramural
Sequencing Center (NISC),
Gaithersburg, Maryland;
Web site: <http://www.nisc.nih.gov/>
Contact: nisc.mgc@nih.gov
Akhter, N., Ayale, K., Beckstrom-Sternberg, S.M., Benjamin, B.,
Blakesley, R.W., Bouffard, G.G., Breen, K., Brinkley, C., Brooks, S.,
Dierich, N.L., Grante, S., Guan, X., Gupta, J., Haghighi, P.,
Hansen, N., Ho, S.-L., Karlins, E., Kwong, P., Laric, P., Legaspi, R.,
Maduro, Q.L., Masiello, C., Maskeri, B., Mastrian, S.D., McCloskey, J.C.,
McDowell, J., Pearson, R., Stantipop, S., Thomas, P.J., Touchman, J.W.,
Tsurgent, C., Vogt, J.B., Walker, M.A., Wetherby, K.D., Wiggins, L.,
Young, A., Zhang, L.-H. and Green, E.D.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: <http://image.lnl.gov>
Series: IRM Plate: 27 Row: m Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 8922941.
Location/Qualifiers

FEATURES
source

1..1503
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:19506 IMAGE:4129973"
/issue_type="Brain, neuroblastoma"
/clone_lib="NIH MGC 19"
/lab_host="DH10B-R"
/note="Vector: pOTB7"
1..1503
/gene="LPAAT-e"
/note="synonym: FLJ11210"
/db_xref="LocusID:55326"
257..1351
/codon_start=1
/product="LPAAT-e protein"
/protein_id="AAH23550.1"
/db_xref="GI:23272120"
/translation="MILSLVHTYSNRYLLPSVLLGTAPTYVLAMGWRLLSAFLPA
RFOALDDRLVCVOSMVLFFPENYTGVOILLYGDLPKNENIYLANHOSIVDMIVA
DILAIRONLIGHRYVLKELIKMLPLVGCYFAOHGIIYKRSKFNKEMRKQSYV
DACTPMYLVTFPEGRYINPEQTKVLSASQAFPAORGLAVLHVLTPRIKATHVADCM
KNVLDAIVDTVYVGEKDDGGRSSPTMTPEFLCKECPKIHIDRIDKDVPEEOEH
MRMLHERFEIKDKMLIEFESPDPERRRRPFKGSVNSKLSIKTLPSMLILSGITAG
MLMTDGRKLYVNTWYITGLGCLMTITKA"
515..856
/note="PLAC: Region: Phosphate acyltransferases"
/db_xref="CD:smart00563"

ORIGIN

Query Match 18.2%; Score 424; DB 9; Length 1503;
Best Local Similarity 99.8%; Pred. No. 6e-230;
Matches 474; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
DB 1 AGCATCCCTGAGCATGATCCGGAGGGCCGGGGTTCCTTGGCCCGCGGAGC 60
QY 1742 AGCATCCCTGAGCATGATCCGGAGGGCCGGGGTTCCTTGGCCCGCGGAGC 1801
DB 1 AGCATCCCTGAGCATGATCCGGAGGGCCGGGGTTCCTTGGCCCGCGGAGC 60
QY 1802 GGGGCGAGGAGCCCGGACTCGCTACCGGGCCCGGGCGGGCGGCGGCGGCGGCT 1861
DB 61 GGGGCGAGGAGCCCGGACTCGCTACCGGGCCCGGGCGGGCGGCGGCGGCGGCT 120
QY 1862 GGGGCGAGGAGCTTGGAGCGGCTGGCGGGCGGGCGGCGGCGGCGGCGGCTGATGGC 1921

DB 121 GGGGCGAGGAGCTTGGAGCGGGGTGGCGGGCGGGCCCGGGCCCGGGCGGATGGC 180
QY 1922 CGCTGCTGCGCGGACTGAGGCGCGGAGCGGGAGCGGAGCGGAGCGGACTGCGTG 1981
DB 181 CGCTGCTGCGCGGACTGAGGCGCGGAGCGGGAGCGGAGCGGAGCGGACTGCGTG 240
QY 1982 CCGCGAGGCTGAGAGAGATGCGGCTGCTCGGAGTCTCAACGATCCATGCGCTACT 2041
DB 241 CCGCGAGGCTGAGAGATGCGGCTGCTCGGAGTCTCAACGATCCATGCGCTACT 300
QY 2042 GCTGCCGAGCGTCTGCTCTGCTGCGGACGCGCCACCTACGTTGGCTGGGGGCTGTG 2101
DB 301 GCTGCCGAGCGTCTGCTCTGCTGCGGACGCGCCACCTACGTTGGCTGGGGGCTGTG 360
QY 2102 GCGGCTGCTCTGCGGCTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCTGTA 2161
DB 361 GCGGCTGCTCTGCGGCTCTGCTGCGGCGGCGGCGGCGGCGGCGGCGGCTGTA 420
QY 2162 CTGCGTCTACGAGCATGAGTCTCTCTTCTTGAATTAACAACCGGGTCCAG 2216
DB 421 CTGCGTCTACGAGCATGAGTCTCTCTTCTTGAATTAACAACCGGGTCCAG 475

RESULT 12
AX268078 3725 bp DNA linear PAT 26-OCT-2001
LOCUS AX268078
DEFINITION Sequence 6 from Patent WO0173051.
ACCESSION AX268078
VERSION AX268078.1 GI:16516593
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
1
Meyers, R.A., Rudolph-Owen, L., Macbeth, K.J., Tsai, F.Y. and
Williamson, M.
46743 and 27417, novel human acyltransferase family members and
uses therefor
Patent: NO 0173051-A 6 04-OCT-2001;
JOURNAL
Millennium Pharmaceuticals, Inc. (US)
FEATURES
source
1..3725
/organism="Homo sapiens"
/mol_type="unassigned DNA"
/db_xref="taxon:9606"
306..1400
/note="unamed protein product"
/codon_start=1
/protein_id="CAD10243.1"
/db_xref="GI:16516594"
/translation="MILSLVHTYSNRYLLPSVLLGTAPTYVLAMGWRLLSAFLPA
RFOALDDRLVCVOSMVLFFPENYTGVOILLYGDLPKNENIYLANHOSIVDMIVA
DILAIRONLIGHRYVLKELIKMLPLVGCYFAOHGIIYKRSKFNKEMRKQSYV
DACTPMYLVTFPEGRYINPEQTKVLSASQAFPAORGLAVLHVLTPRIKATHVADCM
KNVLDAIVDTVYVGEKDDGGRSSPTMTPEFLCKECPKIHIDRIDKDVPEEOEH
MRMLHERFEIKDKMLIEFESPDPERRRRPFKGSVNSKLSIKTLPSMLILSGITAG
MLMTDGRKLYVNTWYITGLGCLMTITKA"

CDS

ORIGIN
Query Match 12.1%; Score 281; DB 6; Length 3725;
Best Local Similarity 99.4%; Pred. No. 3e-148;
Matches 501; Conservative 0; Mismatches 2; Indels 1; Gaps 1;
DB 1714 TGGCGATGGGTAGAGGTGATGGGGGTGTGAGCATCCCTTGAAGCATCCGGAGGGCGG 1773
DB 21 TGGCGATGGGTAGAGGTGATGGGGGTGTGAGCATCCCTTGAAGCATCCGGAGGGCGG 80
QY 1774 GGGGTTCCCTTGGCTTGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCTGCTACCGGCG 1833
DB 81 GGGGTTCCCTTGGCTTGGCGGCGGAGCGGCGGCGGCGGCGGCGGCGGCTGCTACCGGCG 140

QY 2181 GNGCTCTTCTCTGAGATTACACGGGGTCAG 2216
|||||
Db 181 GTGCTCTTCTTTCGAGATTACACGGGGTCAG 216

RESULT 15

BD196514 4682 bp DNA linear PAT 17-JUL-2003
LOCUS BD196514
DEFINITION Prostatic cancer gene.
ACCESSION BD196514
VERSION BD196514.1 GI:33006284
KEYWORDS JP 2002516657-A/103.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
1 (bases 1 to 4682)
AUTHORS Cohen, D., Blumenfeld, M., Chumakov, I. and Bougueleret, L.
TITLE Prostatic cancer gene
JOURNAL Patent: JP 2002516657-A 103 11-JUN-2002;
GENSET

COMMENT OS Homo sapiens (human)
PN JP 2002516657-A/103
PD 11-JUN-2002
PE 22-DEC-1998 JP 200525562
PR 22-DEC-1997 US 08/996306 09-SRP-1998 US 60/099658 PI
DANIEL COHEN, MARTA BLUMENFELD, ILVA CHUMAKOV, LYDIE BOUGUELERET
C12N15/09, C12N15/09, A01K67/027, C07K14/47, C07K16/18, C12N1/15, PC
C12N1/19,
PC C12N1/21, C12N5/10, C12N5/10, C12P21/08, C12Q1/68, G01N33/50 PC
, C12N15/00, C12N5/00,
PC C12N5/00, C12N15/00
CC ATG
CC TGA
CC AATAAA
FH Key location/Qualifiers
FT misc feature 31..33
FT misc feature 301..303
FT polyA_signal 4655..4660.
location/Qualifiers
1..4682
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"

FEATURES
source
ORIGIN

Query Match 9.3%; Score 216; DB 6; Length 4682;
Best Local Similarity 100.0%; Pred. No. 4e-111;
Matches 216; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2001 CTGCTGTCCCTGGTGCCTCCACAGTACTCCATGCGTACCTGTCGCCAGCGTGTCTC 2060
|||||
Db 1 CTGCTGTCCCTGGTGCCTCCACAGTACTCCATGCGTACCTGTCGCCAGCGTGTCTC 60
QY 2061 CTGGGACACGGCGCCACCTACGTTTGCCCTGGGGGCTGTGGCGGCTCTCCGCTTC 2120
|||||
Db 61 CTGGGACACGGCGCCACCTACGTTTGCCCTGGGGGCTGTGGCGGCTCTCCGCTTC 120
QY 2121 CTGGCGCGCGCGCTTCTACCAAGCGCTGAGAGACCGGCTGTACTGCGTCTTACGAGACATG 2180
|||||
Db 121 CTGGCGCGCGCGCTTCTACCAAGCGCTGAGAGACCGGCTGTACTGCGTCTTACGAGACATG 180
QY 2181 GTGCTCTTCTTCTTCCGAGATTACACCGGGGTCCAG 2216
|||||
Db 181 GTGCTCTTCTTCTTCCGAGATTACACCGGGGTCCAG 216

Search completed: May 26, 2004, 10:29:32
Job time: 8839.79 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 06:13:58 ; Search time 5465.68 Seconds
(without alignments)
12697.364 Million cell updates/sec

Title: US-09-901-484A-179_COPY_1_2324
Perfect score: 2324
Sequence: 1 gtggatctgtgctgcgc.....accgcgcgcgcgcgcgcgcac 2324

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0
Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estmu:*
5: em_estcov:*
6: em_estcpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rod:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	455	19.6	1201	13	BX332287 BX332287
2	418	18.0	822	12	B1197695 B1197695
3	369	15.9	1005	9	AL560635 AL560635
4	335	14.4	787	12	BI913991 BI913991

5	330	14.2	663	12	B116158
6	291	12.5	717	12	BG828889
7	290	12.5	313	13	BU429790
8	281	12.1	644	10	AW969608
9	261	11.2	1201	9	AL518744
10	247	10.6	340	10	BE247552
11	227	9.8	567	13	BX473402
12	227	9.8	892	13	BX154303
13	223	9.6	687	12	BM789163
14	221	9.5	1216	13	BU931112
15	217	9.3	643	13	BX507409
16	195	8.4	489	13	BX491999
17	190	8.2	1201	10	AL549745
18	182	7.8	1201	13	BX365137
19	180	7.7	863	12	BG336356
20	180	7.7	1201	9	AL546290
21	177	7.6	575	14	CD687022
22	161	6.9	405	10	AW499989
23	161	6.9	610	13	BX475132
24	161	6.9	903	10	BE513131
25	154	6.6	494	13	BU430723
26	154	6.6	586	9	AL710152
27	144	6.2	704	9	AU138151
28	144	6.2	739	14	CB990373
29	144	6.2	775	14	CB988492
30	144	6.2	792	14	CB960066
31	144	6.2	834	14	CF993793
32	125	5.4	900	10	BP983549
33	117	5.0	933	14	CB960004
34	116	5.0	1406	14	CD358834
35	89	3.8	552	13	BX490286
36	87	3.7	1185	13	BX436436
37	83	3.6	919	14	CB988242
38	59	2.5	329	10	BE245374
39	59	2.5	932	10	BF315750
40	57	2.5	405	13	BU429366
41	48	2.1	699	12	BI824709
42	46	2.0	677	12	BI833979
43	40	1.7	919	14	CB988242
44	40	1.7	934	13	BX374425
45	38	1.6	147	29	AG221710

ALIGNMENTS

RESULT 1
LOCUS BX332287 1201 bp mRNA linear EST 02-MAY-2003
DEFINITION CDNA clone CS0DC010X07 5-PRIME, mRNA sequence.
ACCESSION BX332287
VERSION BX332287.1 GI:30337119
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1201)
Li, W.B., Gruber, C., Jeessee, J. and Polayes, D.
Full-length cDNA libraries and normalization
Unpublished (2001)
Contact: Genoscope
Genoscope - Centre National de Sequencage
BP 191 91006 Evry cedex - France
Email: seqref@genoscope.cns.fr, Web : www.genoscope.cns.fr
Library was constructed by life technologies, a division of
Invitrogen. This sequence belongs to sequence cluster 5026.f For
more information about this cluster, see
http://www.genoscope.cns.fr/
cgl-bin/cluster.cgi?seq=CS0DC010CE04Q0P1c&cluster=5026.f. Contact :
Feng Liang Email : fliang@lifetech.com URL :
http://fulllength.invitrogen.com/Invitrogen Corporation 1600

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniota; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
AUTHORS	1 (bases 1 to 1005)
TITLE	Li, W.B., Gruber, C., Jesse, J., and Polayes, D.
JOURNAL	Full-length cDNA libraries and normalization Unpublished (2001)
COMMENT	On Feb 15, 2001 this sequence version replaced gi:12907288. Contact: Genoscope Genoscope - Centre National de Sequencage BP 191 91006 Evry cedex - France Email: seque@genoscope.cns.fr, Web : www.genoscope.cns.fr Library was constructed by life technologists, a division of Invitrogen. This sequence belongs to sequence cluster 5026.f For more information about this cluster, see http://www.genoscope.cns.fr/ cgi-bin/cluster.cgi?seq=CSDDL003CB01Q1&cluster=5026.f. Contact : Feng Liang Email : fliang@lifetech.com URL : http://fulllength.invitrogen.com/Invitrogen_Corporation_1600 Faraday Avenue Genoscope sequence ID : CSDDL003CB01Q1.
FEATURES	Location/Qualifiers
Source	1..1005 /organism="Homo sapiens" /mol_type="mRNA" /db_xref="taxon:9606" /clone="CSDDL003YD01" /cell_type="B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED" /cell_line="RAMOS CELL LINE" /clone_id="Homo sapiens B CELLS (RAMOS CELL LINE) COT 25-NORMALIZED" /note="1st strand cDNA was primed with a NotI-oligo (dt) primer. Five prime end enriched, double-strand cDNA was digested with Not I and cloned into the Not I and EcoR V sites of the pCMVSPORT 6 vector. Library was normalized."
ORIGIN	
Query Match	15.9%; Score 369; DB 9; Length 1005;
Best Local Similarity	99.4%; Pred. No. 6,4e-166;
Matches	519; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
Db	1803 GCGCAGCAGCCCGCACTCGCTACCCGGCCCGGCGCGCGGCCCATGCGCTG 1862
	157 GCGCAGCAGCCCGCACTCGCTACCCGGCCCGGCGCGCGGCCCATGCGCTG 216
Qy	1863 GGGCGGAGGCTGGGAGGCGGGCGGGCGCGGCCCGCGGCGCGGCTGATTGGCC 1922
Db	217 GGGCGGAGGCTGGGAGGCGGGCGGGCGGGCGGCCCGCGGCGCGGCTGATTGGCC 276
Qy	1923 GCTGCTGGCGCGCACTGAGGCCGGGAGGCGGGCGGGGAGCGCAGCGGAGCTGCTGC 1982
Db	277 GCTGCTGGCGCGCACTGAGGCCGGGAGGCGGGCGGGGAGCGCAGCGGAGCTGCTGC 336
Qy	1983 GCGCGAGCTGAGAGAGATGCTGTGCTCCCTGCTCTCAACAGTACTCATGCGCTACTG 2042
Db	337 GCGCGAGCTGAGAGAGATGCTGTGCTCCCTGCTCTCAACAGTACTCATGCGCTACTG 396
Qy	2043 CTGCCCAAGCTGCTGCTCTGCGGAGCGGGCGCCACTACGATGCTGGCTGGGGGGCTCGG 2102
Db	397 CTGCCCAAGCTGCTGCTCTGCGGAGCGGGCGCCACTACGATGCTGGCTGGGGGGCTCGG 456
Qy	2103 CGGCTGCTCTCGGCTTCTGCGCGCGCGCTTCTACCAAGGCTGGAAGCAGCGGCTGAC 2162
Db	457 CGGCTGCTCTCGGCTTCTGCGCGCGCGCTTCTACCAAGGCTGGAAGCAGCGGCTGAC 516
Qy	2163 TGGCTTTCACAGAGATGCTGCTTTCTTTCAGAAATTACCGGGGCTCAGGTGAGC 2222
Db	517 TGGCTTTCACAGAGATGCTGCTTTCTTTCAGAAATTACCGGGGCTCAGGTGAGC 576
Qy	2223 CGGCTCCGCGCTCCGGGATCTGGGCTCCACCCGAGCTCCGGGGGCGGAGCTTCCGC 2282
Db	577 CGGCTCCGCGCTCCGGGATCTGGGCTCCACCCGAGCTCCGGGGGCGGAGCTTCCGC 636

Qy	2283	TCCCCCAGCTGGCGAGGCTCACCCGCCGCCGCAGAC	2324
Dd	637	TCCCCACACTGGCGAGGCTCACCCGCCGCCGCAGAC	678
RESULT 4	B1913991	787 bp	mRNA linear EST 16-OCT-2001
LOCUS	60318056P1 NIH_MGC_121 Homo sapiens cDNA clone IMAGE:5244761 5'		
DEFINITION	mRNA sequence.		
ACCESSION	B1913991		
VERSION	B1913991.1 GI:16178288		
KEYWORDS	EST.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.		
TITLE	1 (bases 1 to 787)		
JOURNAL	NIH-MGC http://mgs.nci.nih.gov/.		
COMMENT	National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999)		
	Contact: Robert Strausberg, Ph.D.		
	Email: GSABDS-remail.nih.gov		
	Tissue Procurement: Life Technologies, Inc.		
	cDNA Library Preparation: Life Technologies, Inc.		
	DNA Sequencing by: Incyte Genomics, Inc.		
	Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNLML at:		
	http://image.lnl.gov		
	Plate: L1AM1617 row: 1 column: 18		
FEATURES	High quality sequence, stop: 700.		
source	Location/Qualifiers		
	1..787		
	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/clone="IMAGE:5244761"		
	/lab_host="DH10B"		
	/clone_id="NIH_MGC_121"		
	/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 3 fetal brains, female age 20 weeks, female age 24 weeks, and male age 26 weeks. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.7 kb, insert size range 0.7-3.5 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 017. Note: this is a NIH_MGC Library."		
ORIGIN			
Query Match	14.4%; Score 335; DB 12; Length 787;		
Best Local Similarity	100.0%; Pred. No. 1.5e-149;		
Matches 335; Conservative	0; Mismatches 0; Indels 0; Gaps 0;		
Qy	1990	CTGAGAAGATCTCTGTCCCTGTGGTGCTCAACAGTACTCCATCGCTA	2049
Dd	5	CTGAAGAAGTCTCTGTCCCTGTGGTGCTCAACAGTACTCCATCGCTA	64
Qy	2050	GCGTGTGCTCTGTGGCAAGCGGCCCATCTACGTGTGGCTGGGGGGTCTGTGGCGCTGC	2109
Dd	65	GCGTGTGCTCTGTGGCAAGCGGCCCATCTACGTGTGGCTGGGGGGTCTGTGGCGCTGC	124
Qy	2110	TCTCGGCTCTCTGCGCGCGCGCTTACACAAGGCTGAGACCGGCTGTACTGCGCTC	2169
Dd	125	TCTCGGCTCTCTGCGCGCGCGCTTACACAAGGCTGAGACCGGCTGTACTGCGCTC	184
Qy	2170	ACCAGAGATGATCTCTTTCTTTCGAGATTACACCGGGTCCAGGTGAGCGGCTCC	2229
Dd	185	ACCAGAGATGATCTCTTTCTTTCGAGATTACACCGGGTCCAGGTGAGCGGCTCC	244
Qy	2230	CAGTCCGGGATCTCGGGCTCAACCGAGCTCCGAGGAGCGGACCTTCCGCTCCCCA	2289

```

Db      245  CGCTCCGCGGTCTCCGCGCTCCACCCGAGTCCCGGGGGGCGGAGACTCTCTCCGCTCCCA 304
QY      2290  CAGCTGGGAGAGGTACACCGGCGGCGCGGAC 2234
Db      305  CAGCTGGGAGAGGTACACCGGCGGCGGCGGAC 339

RESULT 5
LOCUS   B116158                      663 bp    mRNA    linear    EST 26-JUN-2001
DEFINITION 602866369F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:5015128 5',
            mRNA sequence.
ACCESSION B116158
VERSION   B116158.1 GI:14567059
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 663)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: DCTD/DTP
FEATURES
     source             1..663
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:5015128"
                        /tissue_type="small cell carcinoma"
                        /cell_line="MGC3"
                        /lab_host="DH10B (phage-resistant)"
                        /note="Organ: lung; Vector: pOTB7; Site 1: XhoI; Site 2:
                        EcoRI; cDNA made by oligo-dT priming. Directionally
                        cloned into EcoRI/XhoI sites using the following 5'
                        adaptor: GGCACGAG(G). Size-selected >500bp for average
                        insert size 1.8kb. Library constructed by Ling Hong in
                        the laboratory of Gerald M. Rubin (University of
                        California, Berkeley) using ZAP-cDNA synthesis kit
                        (Stratagene) and Superscript II RT (Life Technologies)."
```

```

QY      1955  GCGCGGAGCGGAGCGGAGCTCGCTGCGCGCGAGCTAGAAGATGCTGTCCCTGT 2014
Db      241  GCGCGGAGCGGAGCGGAGCGGAGCTCGCTGCGCGCGAGCTAGAAGATGCTGTCCCTGT 300
QY      2015  GCTCCACACGTAATTCATGCGCTACCTGCTGCCAGCGTGTGCTCTCTGCGCACGCGCC 2074
Db      301  GCTCCACACGTAATTCATGCGCTACCTGCTGCCAGCGTGTGCTCTCTGCGCACGCGCC 360
QY      2075  CACCTACGTTGTGGCTGGGGGGGTCTGGCGGCTGCTCTCCGCTTCCCGCGCGCTT 2134
Db      361  CACCTACGTTGTGGCTGGGGGGGTCTGGCGGCTGCTCTCCGCTTCCCGCGCGCTT 420
QY      2135  CTACCAAGCGCTGACGACGACGCGCTGTAAGCTGCTACACAGACATGAGTCTTCTT 2194
Db      421  CTACCAAGCGCTGACGACGACGCGCTGTAAGCTGCTACACAGACATGAGTCTTCTT 480
QY      2195  CGAGATTACACCGGCGTCCAG 2216
Db      481  CGAGATTACACCGGCGTCCAG 502

RESULT 6
LOCUS   BG828889                      717 bp    mRNA    linear    EST 22-MAY-2001
DEFINITION 602752872F1 NIH_MGC_17 Homo sapiens cDNA clone IMAGE:4905832 5',
            mRNA sequence.
ACCESSION BG828889
VERSION   BG828889.1 GI:14176476
KEYWORDS  EST.
SOURCE    Homo sapiens (human)
ORGANISM  Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 717)
AUTHORS   NIH-MGC http://mgi.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT   Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Ling Hong/Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LNL at:
            http://image.llnl.gov
            Plate: LNCMI807 row: 0 column: 17
            High quality sequence stop: 681.
FEATURES
     source             1..717
                        /organism="Homo sapiens"
                        /mol_type="mRNA"
                        /db_xref="taxon:9606"
                        /clone="IMAGE:4905832"
                        /tissue_type="rhabdomyosarcoma"
                        /lab_host="DH10B (phage-resistant)"
                        /note="Organ: muscle; Vector: pOTB7; Site 1: EcoRI;
                        Site 2: XhoI; cDNA made by oligo-dT priming.
                        Directionally cloned into EcoRI/XhoI sites using the
                        following 5' adaptor: GGCACGAG(G). Size-selected >500bp
                        for average insert size 1.8kb. Library constructed by
                        Ling Hong in the laboratory of Gerald M. Rubin (University
                        of California, Berkeley) using ZAP-cDNA synthesis kit
                        (Stratagene) and Superscript II RT (Life Technologies)."
```

```

QY      1747  CCTGAGCCATGATCCGCGGAGGCGCGGCTCTTGTTCGCCGCGGAGCGGCG 1806

Query Match      12.5%; Score 291; DB 12; Length 717;
Best Local Similarity 99.8%; Pred. No. 2.3e-128;
Matches 411; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
```

Db	2	CCCTGAGGCATCGATCCGGAGAGGCCGGGGGTTCCCTTGCTTTGCGCCGCGGAGCCGGCGC	61
QY	1807	ACGCGACCCCGCACTCGCTTAACCGGACC CGGCGCGCGCGCCGAGCCCATGCGGCTTGGGG	1866
Db	62	ACGCGACCCCGGCATCGCTCTTACCGGGCCCGGGCGGCGCGCGGCCCATGCGGCTTGGGG	121
QY	1867	CGAGAGCTGGGAGCGGGTGCGCGGGCGCGGCCCGCGGCCGCTGATTGGCCCGCT	1926
Db	122	CGGAGGCTGGAGCGGGTGCGCGGGCGCGCGG - CGGGCCCGGGCGGTGATTGGCGCGCT	180
QY	1227	GCTGGCCGCGCACTGAGGCCCCGGGAGCGCGGAGCGCAGGCGGAGCTGCTGCGGCC	1986
Db	181	GCTGGCCGCGCACTGAGGCCCCGGGAGCGCGGAGCGCAGGCGGAGCTGCTGCGGCC	240
QY	1987	GAGCTGAGAAATGCTGCTGTTCCTGTGGTGCTCCACAGTACTCCATGGGCTAAGCTGCTGC	2046
Db	241	GAGCTGAGAAATGCTGCTGTTCCTGTGGTGCTCCACAGTACTCCATGGGCTAAGCTGCTGC	300
QY	2047	CCAGGCTGTGCTCTGTGGGCAACGGGCCCACTTACGTGTGGGCTGGGGGGGCTTGGGCGC	2106
Db	301	CCAGGCTGTGCTCTGTGGGCAACGGGCCCACTTACGTGTGGGCTGGGGGGGCTTGGGCGC	360
QY	2107	TGCTCTCCGCGCTTCTCTGCGCGCGCGCTTCTTCAAGCGCTGAGACGAGCGCT	2158
Db	361	TGCTCTCCGCGCTTCTCTGCGCGCGCGCTTCTTCAAGCGCTGAGACGAGCGCT	412

RESULT 7					
LOCUS	BU429790	313 bp	mRNA	linear	EST 09-SEP-2002
DEFINITION	UI-HF-BN0-aff-f-05-0-UI.r1 NIH MGC_50 Homo sapiens cDNA clone IMAGE:3066752 5', mRNA sequence.				

ACCESSION	BU429790	
VERSION	BU429790.1	GI:22768277
KEYWORDS	EST.	

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.

REFERENCE	1 (bases 1 to 313)
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/

TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL	Unpublished (1999)

COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cqapbs-r@mail.nih.gov

Eco RI site shown at the beginning of the sequence
Tissue Procurement: Louis M. Staudt, M.D., Ph.D.

CDNA Library Preparation: M.B. Soares Lab
CDNA Library Arrayed by: M.B. Soares Lab

DNA Sequencing by: M.B. Soares Lab
Clone distribution: MGC clone distribution information can be

found through the I.M.A.G.E. Consortium
www-bio.1nl.gov/bbrp/image/image.html

The following repetitive elements were found in this cDNA sequence:

110-233, >GC_rich#Low complexity
Seq primer: M13 Forward.

FEATURES	Location/Qualifiers
source	1. .313

```
/organism="Homo sapiens"  
/mol_type="mRNA"
```

```
/db_xref="taxon:9606"  
/clone="IMAGE:3066752"
```

```
/tissue_type="lymph"  
/cell_type="germinal center B cells"
```

```
/cell_line="MGC85"  
/lab_host="DH10B (LTI)"
```

```

/clone_lib="NIH_MGC_50"
/note="Vector: pT73-Pac; Site 1: NotI; Site 2: Eco RI;

```

Constructed from size fractionated cytoplasmic mRNA (3.5-4.4kb). Directionally cloned. Cells provided by

Louis M. Staudt, Ph.D. Library preparation by Maria de

ORIGIN

Fatima Bonaldo, Ph.D. and M. Bento Soares, Ph.D. "

Query Match	12.5%;	Score 290;	DB 13;	Length 313;
Best Local Similarity	100.0%;	Pred. No. 7e-128;		
Matches 290;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1687 GAGAGCCCCCTGCCCGGACAGGGGGATGTGGCGATGGGTGAGGTCATGGGGGTGTGAGCCT 1746

Db 7 GAGAGCCCCCTGCCCGGACAGGGGGATGTGGCGATGGGTGAGGTCATGGGGGTGTGAGCCT 66

QY	1747	CCCTGAGGCATCGATCCGGGAGGGCCGGGGTTCCCTTGCTTTGCGCCGGAGCGGGCC	1808
Db	67	CCCTGAGGCATCGATCCGGGAGGGCCGGGGTTCCCTTGCTTTGCGCGGGAGCGGGCGC	126

QY	1807	ACGCGAGCCCCGCACTGCGCTACCCGGCCCCGGGCGCGGAGCCATCGCGCTGGGG	1866
Db	127	ACGCGAGCCCCGCACTGCGCTACCCGGCCCCGGGCGCGGAGCCATCGCGCTGGGG	186

QY 1867 CGAGAGCTGGAGCGGTTGCGGGCGCGGCGCCCGGCGGTGATTGGCGGCT 1922

Db 187 CGAGAGCTGGAGCGGTTGCGGGCGCGGCGCCCGGCGGTGATTGGCGGCT 246

QY 1927 GCTGGCCCGACTGAGGCCCGGAGGCGGGAGCCAGCGGAGCT 1976

Db 247 GCTGGCCCGACTGAGGCCCGGAGGCGGGAGCCAGCGGAGCT 296

RESULT 8
AW369608/C

LOCUS	644 bp	mRNA	linear	EST
LOCUS	644 bp	mRNA	linear	EST 01-JUN-2000
DEFINITION	EST31861685	MAGE	reseque	ences, MAGK
ACCESSION	AM9659608			Homo sapiens
VERSION	AM9659608.1			GI:8159452

KEYWORDS: EST.
SOURCE: Homo sapiens (human)
ORGANISM: Homo sapiens
Bukarotia : Metazoa : Chordata : Craniata : Euteleostomi :

REFERENCE
Mammalia; Eutheria; Primates; Catarrhini. Homnidae; Homo.
1 (bases 1 to 644)
Qi, R., Abdenathy, K., Dharwad, S., Gaspard, R., Gay, C.,
Hedger, P., Speed, A., Sharov, V., Lee, N.H., Verman, T. and
Holt, J.R.

TITLE Quackenbush, J. Assessment of gene expression patterns in a model of colon tumor metastasis using a 19,200 element cDNA microarray

COMMENT
Contact: John Quackenbush
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA

Fax: 301 838 0208
Email: johnq@ctigf.org
Plate: 267

FEATURES	Location/Qualifiers
Source	1. .644 /organism="Homo sapiens"

```

//_xref="caxon:9606"
//clone_lib="MAGE_resesquences",MAGK"
/note="Vector: pBluescriptSkM"

```

Query Match	12.1%	Score 281;	DB 10;	Length 644;
Best Local Similarity	99.7%	Pred. No. 1.5e-123;		

	Machines	U/	mismatches	L/	inserts	U/	gaps	
658	ATAATATATATATATATATATATATATATATATGTTAAAGCATTTCGGCATTCCTTTTAAAGT							717

DB 504 AA 432

QY 718 ACAAATATCTTGAAAGGGTTACATATTAAACATTTTACCA CAGCCAAAGGGAGGA 777


```

1. .56 /
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UDXFP686FP.6162"
/dev_stage="adult"
/lab_host="MDH10B"
/clone_lib="666 (synonym: hlc3)"
/notes="Vector: pRiplex; Site_1: stII; Site_2: stIIB

```

Only Match 9.8%; Score 227; DB 13; Length 892;
 Best Local Similarity 100.0%; Pred. No. 1.4e-97;
 Matches 227; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1729 GTCATGGGGGTGTGAGCATCCCTGAGCCATCGATCCGGAGGGCGCGGGGTTCCCTTGCTT 17888
 Db 1 GTATATGGGGGTGTGAGCATCCCTGAGCCATCGATCCGGAGGGCGCGGGGTTCCCTTGCTT 60

QY	1789		GGCCGCGGGAGCGCGCGAGCAGCCCCCGACACTGCCTAACCCGGGCCCCGGGGCGGGCGC	1848
Db	61		TGCGCGGGGAGCGCGCAGCGAGCCCCCGCACTGCCTAACCCGGGCCCCGGGGCGGGCGC	120
QY	1849		GGCCCATCGGCTGGGGGCGGAGGCTGGGAGACGGGTTGGCGGACGGCGCGGCGCCGAGCCCG	1908
Db	121		GGCCCATCGGCTGGGGGCGGAGGCTGGGAGACGGGTTGGCGGACGGGCGGCGCGGCGCCGAGCCCG	180
QY	1909		GGCGGTGATTGGCGCGCTGCTGGCCCGCAGCTGAGGCGCCGGAGGCGG	1955
Db	181		GGCGGTGATTGGCGCGCTGCTGGCCCGCAGCTGAGGCGCCGGAGGCGG	227
RESULT 13				
BM789163				
LOCUS	BM789163	687 bp	mRNA	linear
DEFINITION	K-EST0068703 S11SNUI Homo sapiens cDNA clone S11SNUI-25-E07 5', mRNA sequence.			
ACCESSION	BM789163			
VERSION	BM789163.1	GI:19137395		
KEYWORDS	EST.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Kim,N.S., Hahn,Y., Oh,J.H., Lee,J.Y., Ahn,H.Y., Chu,M.Y., Kim,M.R.,			

TITLE 21C Frontier Korean EST Project 2001
JOURNAL Unpublished (2002)
COMMENT Contact: Kim YS
Genome Research Center
Korea Research Institute of Bioscience & Biotechnology
52 Boeun-dong Yuseong-gu, Daejeon 305-333, South Korea
Tel: +82-42-860-4470
Fax: +82-42-860-4409
Email: yongsung@mail.kribb.re.kr
Plate: 25 row: E column: 07
High quality sequence stop: 687.
location/Qualifiers
1. 687

```

/organism="Homo sapiens"
/mol_type="rRNA"
/db_xref="taxon:9606"
/clone="S11SN01-25-E07"
/sex="M"
/tissue_type="Stomach"
/cell_type="Lymphoblast-like"
/cell_line="SNu-1"
/lab_host="Top10F"
/clone_idb="S11SN01"
/notes="Organ: Stomach; Vector: pME18-FL3; Site 1: XhoI; Site 2: XhoI; The poly (A) + RNA was dephosphorylated with bacterial alkaline phosphatase (BAP) and then decapped with tobacco acid pyrophosphatase (TAP). The decapped intact mRNA was ligated with DNA-RNA linker including SfiI site by treatment of T4 RNA ligase and the first strand cDNA was synthesized with Superscript II using SfiI oligo-dT primer. After first strand synthesis, RNA was degraded by NaOH treatment and cDNA was amplified by PCR reaction. The PCR products were digested with SfiI and cloned into DraIII- digested pME185-FL3 vector. The obtained cDNA vectors were used for transformation of competent cells E. coli Top10F by electroporation method. The cDNA libraries constructed by this method are full-length enriched cDNA library."

```

```

ORIGIN
Query Match          9.6%; Score 223; DB 12; Length 687;
Best Local Similarity 100.0%; Pred. No. 1.2e-95;
Matches 223; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY	1299	GAAGATGCTGTCGTCCTCGGAGGCGCCACAGATATCATAGCGCTACCTGGCTGCCACAGT	2053
Db	1	GAAATGCTGCTGTCCTCGGAGGCTCCACAGATATCATAGCGCTACCTGGCTGCCACAGT	60
QY	2054	CGTGCCTCTGGGACAGGAGCGCCACTTACGTGTGGCTGGGGGGTCTGGCGCGCTGCTC	2113
Db	61	CGTGCCTCTGGGACAGGAGCGCCACTTACGTGTGGCTGGGGGGTCTGGCGCGCTGCTC	120
QY	2114	CGCCTTCTGCGCGCCGCGCTTCTACCAAGGCGCTGACAGACCGGCTGTACTGCGCTTACCA	2173
Db	121	CGCCTTCTGCGCGCCGCGCTTCTACCAAGGCGCTGACAGACCGGCTGTACTGCGCTTACCA	180
QY	2174	GAGCATGCTGCTCTTCTTCTTCAGAAATTACACCGGGGTCCAG	2216
Db	181	GAGCATGCTGCTCTTCTTCTTCAGAAATTACACCGGGGTCCAG	223

RESULT 14	
LOCUS	BU931112
DEFINITION	BU931112 1216 bp mRNA linear EST 18-OCT-2002
ACCESSION	AGNCOURT_10475891 NIH_MGC_109 Homo sapiens cDNA clone
VERSION	IMAGE:6671220 5', mRNA sequence.
KEYWORDS	BU931112
SOURCE	BU931112.1 GI:24119931
ORGANISM	EST.
AUTHORS	Homo sapiens (human)
TITLE	Homo sapiens
JOURNAL	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
COMMENT	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 1216) NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished (1999) Contact: Robert Strausberg, Ph.D.

Email: cgaps-i@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Rubin Laboratory
DNA Sequencing by: The I.M.A.G.E. Consortium (LNL)
Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://image.llnl.gov>
Plate: L10CM2949 row: m column: 12
High quality sequence stop: 269.
Location/Qualifiers
1. .1216

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:6671220"
/tissue_type="teratocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/clone_1b="NIH MGC 103"
/notes="organ: ovary; Vector: POTB1; Site 1: EcoRI; Site 2: XhoI; cDNA made by oligo-dT priming. Directionally cloned into EcoRI/XhoI sites using the following 5' adaptor: GGCGCGAG(G). Library constructed by Ling Hong in the laboratory of Gerald M. Rubin (University of California, Berkeley) using ZAP-cDNA synthesis kit (Stratagene) and Superscript II RT (Life Technologies). Note: this is a NIH MGC library."

```

ORIGIN	Query Match	9.5%	Score 221	DB 13	Length 1216
	Best Local Similarity	100.0%	Pred. No. 1.1e-94		
	Matches 221	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1739	GTGAGCATCCCTGAGCATGATCCGGAGGCGCGGAGTTCCTTGCTTTGCGCGCGG	1798		
Db	34	GTGAGCATCCCTGAGCATGATCCGGAGGCGCGGAGTTCCTTGCTTTGCGCGCGG	93		
QY	1799	AGCGGCGACGACCCCGCATCTGCGCTACCCGGACCCGGAGCGGCGCGGCCCATATCG	1858		

Accession	Sequence	Position
Db	AGCGGCGGACCGCAGCCTCGCTCACTCCGGGCCGGGGCGGGGCCCATCG	153
QY	GCTGGGGGCGGAGGCTGGAGCCGGGTGGCGGCGCGGCCCGGGCCGGGTGATT	1918
Db	GCTGGGGGCGGAGGCTGGAGCCGGGTGGCGGCGCGGCCCGGGCCGGGTGATT	213
QY	GGCGGCGTGGTGGCGGCGGCTGAGGCGCGGGAGGCGGGCGG	1959
Db	GGCGGCGTGGTGGCGGCGGCTGAGGCGCGGGAGGCGGGCGG	254

RESULT 15

LOCUS	BX5074409	643 bp	mrna	linear	EST 04-SEP-2003
DEFINITION	DKEZ57790084_r1_779 (synonym: hnccl1) Homo sapiens CDNA clone				
	DKEZ57790084_5, mRNA sequence.				

ACCESSION

KEYWORDS

SOURCE
ORGANIS

REFERENCE

100

TITTLE

ТОПНАТ

COMMENT

INGOLSTAEDTER LANDSTR. 1, D-85764 NEUMUENBERG, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
Heidelberg/Germany) within the cDNA sequencing consortium of the
German Genome Project.
No 91 sequence available.
This clone (DKFZp79O084) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg; GERMAN; Email: clone@rzd.de.
Location/Accession

sour

ORIGIN

Query match	9.58	Score 217	66.13	Length 852
Best Local Similarity	100.0%	Pred. No.	9.4e-93	
Matches 217; Conservative	0;	Mismatches	0;	Indels 0; Gaps 0;

Qy	2000	GCTGCTGTCCTGGAGCTCCACAGCTATCTCATGAGGCTACCGCTGCCAAGGTGTGT	205
Db	2	GCTGCTGTCCTGGAGCTCCACAGCTATCTCATGAGGCTACCGCTGCCAAGGTGTGT	61
Qy	2060	CCTGGGACAGGCGCCACCTAGTATGTTGAGCTTGAGCGAGCTGTCCGCTT	211
Db	62	CCTGGGACAGGCGCCACCTAGTATGTTGAGCTTGAGCGAGCTGTCCGCTT	121
Qy	2120	CCTGCGCGCCCGCTTCTACCAAGCGCTGAGCGAGCTGTACTGCTCTACCAAGAGT	217
Db	122	CCTGCGCGCCCGCTTCTACCAAGCGCTGAGCGAGCTGTACTGCTCTACCAAGAGT	181

Oy 2180 GGTGCTTTCCTTCGAGATTACCGGGGTTCCAG 221
| | | | | | | | | | | | | | | | | |
Db 182 GGTCCTCCTTCCTCGAATTACCGGGGTTCCAG 218

Search completed: May 26, 2004, 12:51:02
Job time : 5471.68 secs

THIS PAGE LEFT BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 26, 2004, 06:13:58 ; Search time 2582.32 Seconds
(without alignments)
12697.364 Million cell updates/sec

Title: US-09-901-484A-179_COPY_3899_4996

Perfect score: 1098
Sequence: 1 gtctctgtctccgcgattt.....agacttaagtcctcgat 1098

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 27513289 seqs, 14931090276 residues

Word size : 0

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2.1	2.1	209	13	BQ989814
2	2.1	2.1	210	13	BQ847561
3	2.1	2.1	210	13	BQ857992
4	2.1	2.22	13		BQ992394

5	2.1	2.1	223	13	BQ847396
6	2.1	2.1	259	13	BQ013709
7	2.1	2.1	353	13	BQ981419
8	2.1	432	9		AV665533
9	2.1	551	13		BU011575
10	2.1	607	9		AV665534
11	2.1	921	29		CNS0200J
12	2.1	921	29		CNS0200J
13	2.0	367	28		AV334716
14	2.0	367	28		AZ726944
15	2.0	424	13		BQ597016
16	2.0	466	9		AI431315
17	2.0	498	28		BH742910
18	2.0	660	9		AI063012
19	2.0	663	28		BH450006
20	2.0	668	29		CE279192
21	2.0	682	14		CA993731
22	2.0	686	29		AG144022
23	2.0	726	28		BZ508298
24	2.0	786	28		BZ070860
25	2.0	1757	11		AK013575
26	1.9	224	14		CP426180
27	1.9	241	9		AV254417
28	1.9	283	12		BI288613
29	1.9	283	12		BI288613
30	1.9	284	10		BB244365
31	1.9	292	9		AI435529
32	1.9	296	28		BZ859441
33	1.9	338	10		AW434852
34	1.9	368	13		BQ492779
35	1.9	430	12		BM106873
36	1.9	436	29		CE610771
37	1.9	475	21		CC557326
38	1.9	481	13		BX721433
39	1.9	492	28		BZ845349
40	1.9	516	12		BI815360
41	1.9	536	13		BQ521596
42	1.9	544	14		WS2378
43	1.9	545	28		BH930122
44	1.9	579	12		BU542168
45	1.9	581	12		BI586329

ALIGNMENTS

RESULT 1
BQ989814
LOCUS
DEFINITION OGR18L11.yg.ab1 OG_EFGHJ lettuce serritola Lactuca sativa cDNA clone
ACCESSION OGR18L11, mRNA sequence.
VERSION BQ989814.1 GI:22409349
KEYWORDS
SOURCE
ORGANISM
Lactuca sativa

REFERENCE
AUTHORS
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L., Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J., Ellison, P., Kolkmann, J., Slabaugh, M.S., Livingston, K., Zhou, Y., Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compgenome.ucdavis.edu/
Unpublished (2002)

CONTACT: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Aamundson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742

Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig QG_CA_Contig7771, see <http://cgpdb.ucdavis.edu/>
 for details.
 Plate: QGF18 row: L column: 11.
 Location/Qualifiers
 1..209

/organism="Lactuca sativa"
 /mol_type="mRNA"
 /cultivar="L.serricola"
 /db_xref="taxon:4236"
 /clone="QGf18L11"
 /lab_host="E.coli"
 /clone_lib="QG_EFGHJ lettuce serricola"
 /note="Vector: pBRCDNASf1AB. The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>
 TAG_TISSUE=roots
 TAG_LIB=QG_EFGHJ lettuce serricola
 TAG_SEQ=GTTCACGGG"

ORIGIN

Query Match 2.1%; Score 23; DB 13; Length 209;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 TTTATTAATTATATATTATAA 502
 ||||||||||||||||||||
 Db 174 TTTATTAATTATATATTATAA 196

RESULT 2
 BQ847561 210 bp mRNA linear EST 14-AUG-2002
 LOCUS QGA5909.yg.ab1 QG ABCDI lettuce salinas Lactuca sativa cDNA clone
 DEFINITION BQ847561 mRNA sequence.
 ACCESSION BQ847561
 VERSION BQ847561.1 GI:22231980
 KEYWORDS EST.
 SOURCE Lactuca sativa
 ORGANISM Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asterales; Cichorioideae; Cichorioideae; Lactuca.
 1 (bases 1 to 210)
 Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.
 Lettuce and Sunflower ESTs from the Composite Genome Project
<http://compgenomics.ucdavis.edu/>
 Unpublished (2002)
 CONTACT: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore lab
 University of California at Davis (UCD)
 Asmundo Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig QG_CA_Contig7771, see <http://cgpdb.ucdavis.edu/>
 for details.
 Plate: QGA3 row: 9 column: 09.
 Location/Qualifiers
 1..210
 /organism="Lactuca sativa"
 /mol_type="mRNA"

/cultivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QGA3909"
 /lab_host="E.coli"
 /clone_lib="QG_ABCDI lettuce salinas"
 /note="Vector: pBRCDNASf1AB. The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpdb.ucdavis.edu/>
 TAG_TISSUE=roots
 TAG_LIB=QG ABCDI lettuce salinas
 TAG_SEQ=GTTCACGGG"

ORIGIN

Query Match 2.1%; Score 23; DB 13; Length 210;
 Best Local Similarity 100.0%; Pred. No. 24;
 Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 TTTATTAATTATATATTATAA 502
 ||||||||||||||||||||
 Db 174 TTTATTAATTATATATTATAA 196

RESULT 3
 BQ857992 210 bp mRNA linear EST 14-AUG-2002
 LOCUS QGB9A22.yg.ab1 QG ABCDI lettuce salinas Lactuca sativa cDNA clone
 DEFINITION QGB9A22, mRNA sequence.
 ACCESSION BQ857992
 VERSION BQ857992.1 GI:22243457
 KEYWORDS EST.
 SOURCE Lactuca sativa
 ORGANISM Lactuca sativa
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids; campanulids; Asterales; Asterales; Cichorioideae; Cichorioideae; Lactuca.
 1 (bases 1 to 210)
 Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Rieseberg,L., Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J., Ellison,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y., Lai,Z., Church,S., Jackson,L. and Bradford,K.
 Lettuce and Sunflower ESTs from the Composite Genome Project
<http://compgenomics.ucdavis.edu/>
 Unpublished (2002)
 CONTACT: Alexander Kozik [R.W.Michelmore]
 Department of Vegetable Crops, R.W.Michelmore lab
 University of California at Davis (UCD)
 Asmundo Hall, UCD, Davis, CA 95616, USA
 Tel: 1-(530)-742-1742
 Fax: 1-(530)-752-9659
 Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
 belongs to contig QG_CA_Contig7771, see <http://cgpdb.ucdavis.edu/>
 for details.
 Plate: QGB3 row: A column: 22.
 Location/Qualifiers
 1..210
 /organism="Lactuca sativa"
 /mol_type="mRNA"
 /cultivar="Salinas"
 /db_xref="taxon:4236"
 /clone="QGB9A22"
 /lab_host="E.coli"
 /clone_lib="QG ABCDI lettuce salinas"
 /note="Vector: pBRCDNASf1AB. The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each

source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpb.ucdavis.edu/>
TAG LIB=OG ABCDI lettuce salinas
TAG_SEQ=GTGCACGGG"

Query Match 2.1%; Score 23; DB 13; Length 210;
Best Local Similarity 100.0%; Pred. No. 24;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 TTTATTAATTTATATATATTTAA 502
|||||
174 TTTATTAATTTATATATATTTAA 196

RESULT 4
BQ992394 222 bp mRNA linear EST 21-AUG-2002
LOCUS QG25103.yg.ab1 QG_EFGHU lettuce serriola Lactuca sativa cDNA clone
DEFINITION QG25103, mRNA sequence.
ACCESSION BQ992394
VERSION BQ992394.1 GI:22411929
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asterales; Cichorioideae;
Cichorieae; Lactuca.

REFERENCE 1 (bases 1 to 222)

AUTHORS

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
<http://compenomics.ucdavis.edu/>
Unpublished (2002)

JOURNAL

COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmndson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig7771, see <http://cgpb.ucdavis.edu/>
for details.
Plate: QG25 row: 1 column: 03.

FEATURES
source Location/Qualifiers
1..222

/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.seriola"
/db_xref="taxon:4236"
/clone="QG25103"
/lab_host="E.coli"
/clone_lib="QG_EFGHU lettuce serriola"
/note="Vector: pBRCDNASf1AB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpb.ucdavis.edu/>
TAG_TISSUE=roots
TAG_LIB=OG_EFGHU lettuce serriola
TAG_SEQ=GTGCACGGG"

ORIGIN

Query Match 2.1%; Score 23; DB 13; Length 222;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 TTTATTAATTTATATATATTTAA 502
|||||
174 TTTATTAATTTATATATATTTAA 196

RESULT 5
BQ847396

LOCUS QGA2a06.yg.ab1 QG ABCDI lettuce salinas Lactuca sativa cDNA clone
DEFINITION QGA2a06, mRNA sequence.
ACCESSION BQ847396
VERSION BQ847396.1 GI:22231616
KEYWORDS EST.

SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asterales; Cichorioideae;
Cichorieae; Lactuca.

REFERENCE 1 (bases 1 to 223)

AUTHORS

Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Rieseberg, L.,
Lin, H., van Damme, M., Lavelle, D., Chevalier, P., Ziegler, J.,
Ellison, P., Kolman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
<http://compenomics.ucdavis.edu/>
Unpublished (2002)

JOURNAL

COMMENT Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Asmndson Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Contig7771, see <http://cgpb.ucdavis.edu/>
for details.
Plate: QGA2 row: a column: 06.

FEATURES
source Location/Qualifiers
1..223

/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="Salinas"
/db_xref="taxon:4236"
/clone="QGA2a06"
/lab_host="E.coli"
/clone_lib="QG_ABCDI lettuce salinas"
/note="Vector: pBRCDNASf1AB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformations made with four size classes to minimize size bias. Details of each source of RNA and library construction can be obtained at <http://cgpb.ucdavis.edu/>
TAG_TISSUE=roots
TAG_LIB=OG ABCDI lettuce salinas
TAG_SEQ=GTGCACGGG"

ORIGIN

Query Match 2.1%; Score 23; DB 13; Length 223;
Best Local Similarity 100.0%; Pred. No. 23;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 480 TTTATTAATTTATATATATTTAA 502
|||||
174 TTTATTAATTTATATATATTTAA 196

ORIGIN

RESULT 6
BU013709

LOCUS
DEFINITION BU013709 259 bp mRNA linear EST 22-AUG-2002
QG5G07.yg.ab1 QG_EFGHU lettuce serriola Lactuca sativa cDNA clone
ACCESSION
VERSION BU013709
KEYWORDS
SOURCE Lactuca sativa
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 259)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Riesberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
Ellison,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
Lai,Z., Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Composite Genome Project
<http://comgenomics.ucdavis.edu/>
Unpublished (2002)
JOURNAL
COMMENT
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Assumdsen Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Config7771, see <http://cgpdb.ucdavis.edu/>
for details.
Plate: QG5 row: G column: 07.

FEATURES
source
1..259
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QG5G07"
/lab_host="E.coli"
/clone_lib="QG_EFGHU lettuce serriola"
/note="Vector: pBRCDNA5flab; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at <http://cgpdb.ucdavis.edu/>
TAG TISSUE=roots
TAG_LIB=QG_EFGHU lettuce serriola
TAG_SEQ=GTGACGCGG"

ORIGIN
Query Match 2.1%; Score 23; DB 13; Length 259;
Best Local Similarity 100.0%; Pred. No. 22;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 TTTATTATTATATATATATATAA 502
DB 175 TTTATTATTATATATATATAA 197

RESULT 7
BO981419
LOCUS
DEFINITION BO981419 353 bp mRNA linear EST 21-AUG-2002
QGE13115.yg.ab1 QG_EFGHU lettuce serriola Lactuca sativa cDNA clone
ACCESSION
VERSION BO981419
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
1 (bases 1 to 432)
Takasuga,A., Hirotsune,S., Itoh,R., Jitchazon,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using

SOURCE
ORGANISM
Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichoriaceae; Lactuca.
1 (bases 1 to 353)
Kozik,A., Michelmore,R.W., Knapp,S., Matvienko,M., Riesberg,L.,
Lin,H., van Damme,M., Lavelle,D., Chevalier,P., Ziegler,J.,
Ellison,P., Kolman,J., Slabaugh,M.S., Livingston,K., Zhou,Y.,
Lai,Z., Church,S., Jackson,L. and Bradford,K.
Lettuce and Sunflower ESTs from the Composite Genome Project
<http://comgenomics.ucdavis.edu/>
Unpublished (2002)
JOURNAL
COMMENT
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Assumdsen Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig QG_CA_Config7771, see <http://cgpdb.ucdavis.edu/>
for details.
Plate: QGE13 row: L column: 15.

FEATURES
source
1..353
/organism="Lactuca sativa"
/mol_type="mRNA"
/cultivar="L.serriola"
/db_xref="taxon:4236"
/clone="QGE13115"
/lab_host="E.coli"
/clone_lib="QG_EFGHU lettuce serriola"
/note="Vector: pBRCDNA5flab; The library was constructed
from 10 different sources of RNA from a single genotype.
Separate cDNAs were generated using primers that
incorporated unique 5' and 3' tags to distinguish each
source of RNA. cDNAs were then pooled, size-fractionated,
directionally cloned into a custom medium-copy vector and
transformations made with four size classes to minimize
size bias. Details of each source of RNA and library
construction can be obtained at <http://cgpdb.ucdavis.edu/>
TAG TISSUE=flowers environmental stress
TAG_LIB=QG_EFGHU lettuce serriola
TAG_SEQ=CGAATCGCGG"

ORIGIN
Query Match 2.1%; Score 23; DB 13; Length 353;
Best Local Similarity 100.0%; Pred. No. 21;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 TTTATTATTATATATATATATAA 502
DB 104 TTTATTATTATATATATATAA 126

RESULT 8
AV665533
LOCUS
DEFINITION AV665533 432 bp mRNA linear EST 28-NOV-2001
AV665533 Bos taurus brain fetus Bos taurus cDNA clone E1BR049B04
3', mRNA sequence.
ACCESSION
VERSION AV665533
KEYWORDS
SOURCE
ORGANISM
Bos taurus (cow)
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 432)
Takasuga,A., Hirotsune,S., Itoh,R., Jitchazon,A., Suzuki,H., Aso,H.
and Sugimoto,Y.
Establishment of a high throughput EST sequencing system using

poly(A) tail-removed cDNA libraries and determination of 36,000 bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
JOURNAL 21570554
MEDLINE 11713328
PUBMED
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@ccococ.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
Location/Qualifiers

FEATURES
source 1..432
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1BR049B04"
/tissue_type="brain"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus brain fetus"
/note="Vector: pZLI; Site 1: SalI; Site 2: NotI; Poly A was deleted from a NotI site"

ORIGIN

Query Match 2.1%; Score 23; DB 9; Length 432;
Best Local Similarity 100.0%; Pred. No. 19;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 480 TTTATTAATTTATATATATTTAA 502
|||||
Db 350 TTTATTAATTTATATATTTAA 372

RESULT 9
BU011575 551 bp mRNA linear EST 22-AUG-2002
LOCUS OG16071.YG.ab1 OG_EFGH lettuce serritola lactuca sativa cDNA clone
DEFINITION OG16071.YG.ab1 OG_EFGH lettuce serritola lactuca sativa cDNA clone
ACCESSION BU011575
VERSION BU011575.1 GI:22445970
KEYWORDS EST.
SOURCE Lactuca sativa
ORGANISM Lactuca sativa
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
asterids; campanulids; Asterales; Asteraceae; Cichorioideae;
Cichorieae; Lactuca.
1 (bases 1 to 551)
Kozik, A., Michelmore, R.W., Knapp, S., Matvienko, M., Riesberg, L.,
Lin, H., Van Damme, M., Lavelle, D., Chevaller, P., Ziegler, J.,
Ellison, P., Koltman, J., Slabaugh, M.S., Livingston, K., Zhou, Y.,
Lai, Z., Church, S., Jackson, L. and Bradford, K.
Lettuce and Sunflower ESTs from the Compositae Genome Project
http://compositae.genomics.ucdavis.edu/
Unpublished (2002)
Contact: Alexander Kozik [R.W.Michelmore]
Department of Vegetable Crops, R.W.Michelmore Lab
University of California at Davis (UCD)
Assumend Hall, UCD, Davis, CA 95616, USA
Tel: 1-(530)-742-1742
Fax: 1-(530)-752-9659
Email: akozik@atgc.org [michelmore@vegmail.ucdavis.edu]
belongs to contig OG_CA_Contig777, see http://cgpdb.ucdavis.edu/
for details.
Plate: OG16 row: J column: 21.
Location/Qualifiers
1..551
/organism="Lactuca sativa"
/mol_type="mRNA"

/cultivar="L. serritola"
/db_xref="taxon:4236"
/clone="OG16071"
/lab_host="E.coli"
/clone_lib="OG_EFGH lettuce serritola"
/note="Vector: pRCMDVAST1AB; The library was constructed from 10 different sources of RNA from a single genotype. Separate cDNAs were generated using primers that incorporated unique 5' and 3' tags to distinguish each source of RNA. cDNAs were then pooled, size-fractionated, directionally cloned into a custom medium-copy vector and transformed. Details of each source of RNA and library construction can be obtained at http://cgpdb.ucdavis.edu/
TAG_TISSUE=flowers pre-fertilized
TAG_LIB=OG_EFGH lettuce serritola
TAG_SEQ=GCTTACGCG3"

ORIGIN

Query Match 2.1%; Score 23; DB 13; Length 551;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 480 TTTATTAATTTATATATTTAA 502
|||||
Db 372 TTTATTAATTTATATATTTAA 394

RESULT 10
AV665534 607 bp mRNA linear EST 28-NOV-2001
LOCUS AV665534 Bos taurus brain fetus Bos taurus cDNA clone E1BR049B04
DEFINITION AV665534 5' mRNA sequence.
ACCESSION AV665534
VERSION AV665534.1 GI:3924564
KEYWORDS EST.
SOURCE Bos taurus (cow)
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 607)
Takasuga, A., Hirotsune, S., Itoh, R., Jitohzono, A., Suzuki, H., Aso, H.
and Sugimoto, Y.
Establishment of a high throughput EST sequencing system using
poly(A) tail-removed cDNA libraries and determination of 36,000
bovine ESTs
Nucleic Acids Res. 29 (22), E108 (2001)
JOURNAL 21570554
MEDLINE 11713328
PUBMED
COMMENT Contact: Yoshikazu Sugimoto
Animal Genetics Division
Shirakawa Institute of Animal Genetics
Odakura, Nishigo, Nishi-shirakawa, Fukushima 961-8061, Japan
Tel: 81-248-25-5641
Fax: 81-248-25-5725
Email: kazusugi@ccococ.ocn.ne.jp
Single pass sequencing.
This clone was obtained from a polyA-deleted cDNA library.
Location/Qualifiers
1..607
/organism="Bos taurus"
/mol_type="mRNA"
/db_xref="taxon:9913"
/clone="E1BR049B04"
/tissue_type="brain"
/dev_stage="fetus"
/lab_host="DH10B"
/clone_lib="Bos taurus brain fetus"
/note="Vector: pZLI; Site 1: SalI; Site 2: NotI; Poly A was deleted from a NotI site"

Query Match 2.1%; Score 23; DB 9; Length 607;
Best Local Similarity 100.0%; Pred. No. 18;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 480 TTTATTATTATATATATATTA 502
|||||
451 TTTATTATTATATATATATTA 429

Db

RESULT 11
CNS0300J/c 771 bp DNA linear GSS 01-SEP-2000
LOCUS Tetradon nigroviridis genome survey sequence PUC-Orl end of clone
DEFINITION 203123 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL222508.1 GI:7881327
VERSION GSS; genome survey sequence.
KEYWORDS Tetradon nigroviridis
SOURCE Tetradon nigroviridis
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetradon.

REFERENCE 1
AUTHORS Roest Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fitzmes, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE Estimate of human gene number provided by genome-wide analysis
using Tetradon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645

REFERENCE 2
AUTHORS Roest Crolius, H., Jallion, O., Dasilva, C., Ozouf-Costaz, C.,
Fitzmes, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143

REFERENCE 3
AUTHORS (bases 1 to 771)
TITLE Genoscope.
DIRECT SUBMISSION Direct Submission
SUBMITTED (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.
location/Qualifiers
1.771
/organism="Tetradon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="203123"
/clone_1lb="G"
/note="Genoscope sequence ID : C0AG203AE12SP1-end :
PUC-Orl"

ORIGIN
Query Match 2.1%; Score 23; DB 29; Length 771;
Best Local Similarity 100.0%; Pred. No. 17;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 TCTTTTCTTGAAATGTTTT 363
|||||
545 TCTTTTCTTGAAATGTTTT 523

Db

RESULT 12

CNS02TZP 921 bp DNA linear GSS 01-SEP-2000
LOCUS Tetradon nigroviridis genome survey sequence T7 end of clone
DEFINITION 166C04 of library G from Tetradon nigroviridis, genomic survey
sequence.
ACCESSION AL213828.1 GI:7872647
VERSION GSS; genome survey sequence.
KEYWORDS Tetradon nigroviridis
SOURCE Tetradon nigroviridis
ORGANISM Tetradon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodontidae; Tetradon.

REFERENCE 1
AUTHORS Roest Crolius, H., Jallion, O., Dasilva, C., Bouneau, L., Fisher, C.,
Bernot, A., Fitzmes, C., Wincker, P., Brottier, P., Quetier, F.,
Saurin, W. and Weissenbach, J.
TITLE Estimate of human gene number provided by genome-wide analysis
using Tetradon nigroviridis DNA sequence
JOURNAL Nat. Genet. 25 (2), 235-238 (2000)
MEDLINE 20296633
PUBMED 10835645

REFERENCE 2
AUTHORS Roest Crolius, H., Jallion, O., Dasilva, C., Ozouf-Costaz, C.,
Fitzmes, C., Fischer, C., Bouneau, L., Billault, A., Quetier, F.,
Saurin, W., Bernot, A. and Weissenbach, J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
JOURNAL Genome Res. 10 (7), 939-949 (2000)
MEDLINE 20359837
PUBMED 10899143

REFERENCE 3
AUTHORS (bases 1 to 921)
TITLE Genoscope.
DIRECT SUBMISSION Direct Submission
SUBMITTED (12-APR-2000) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
Web : www.genoscope.cns.fr)
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetradon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.
location/Qualifiers
1.921
/organism="Tetradon nigroviridis"
/mol_type="genomic DNA"
/db_xref="taxon:99883"
/clone="166C04"
/clone_1lb="G"
/note="Genoscope sequence ID : C0AG166BB02LP1-end : T7"

ORIGIN
Query Match 2.1%; Score 23; DB 29; Length 921;
Best Local Similarity 100.0%; Pred. No. 16;
Matches 23; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 341 TCTTTTCTTGAAATGTTTT 363
|||||
697 TCTTTTCTTGAAATGTTTT 719

Db

RESULT 13
AV334716 236 bp mRNA linear EST 11-NOV-1999
LOCUS AV334716 RIKEN full-length enriched, adult male medulla oblongata
DEFINITION Mus musculus cDNA clone 6330563N19 3', mRNA sequence.
ACCESSION AV334716.1 GI:6374768
VERSION EST.
KEYWORDS Mus musculus (house mouse)
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

SOURCE Plasmodium falciparum 3D7
 ORGANISM Plasmodium falciparum 3D7
 Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 REFERENCE 1 (bases 1 to 424)
 AUTHORS Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Page,D.,
 Marra,M., Hillier,L., Martin,J., Wylie,T., Dante,M., Theising,B.,
 Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,
 Tsagarisvilli,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,
 Maguire,L., Richey,J., Madkins,J., Kennedy,S., Levine,D.,
 Waterston,R., Wilson,R. and Sibley,D.
 TITLE Washu Plasmodium EST Project
 JOURNAL Unpublished (2001)
 COMMENT Contact: L. David Sibley
 Washu Plasmodium EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
 Tel: 314 286 1800
 Fax: 314 286 1810
 Email: est@watscn.wustl.edu
 Library was constructed by Debopam Chakrabarti DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: L. David Sibley
 (sibley@borcim.wustl.edu), Washington University
 Seq primer: -40UP from Gibco
 High quality sequence stop: 423.
 Location/Qualifiers

FEATURES

source
 1..424
 /organism="Plasmodium falciparum 3D7"
 /mol_type="mRNA"
 /db_xref="taxon:36329"
 /lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
 /clone_lib="Plasmodium falciparum 3D7 asexual cDNA"
 /note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2:
 XhoI; Library was constructed by Debopam Chakrabarti.
 Total RNA samples were isolated from mixed stage
 saponin(0.1%) lysed P. falciparum 3D7 infected
 erythrocytes by the acidic guanidinium-phenol chloroform
 method. The poly A+ RNA was isolated by the polyAT-Tract
 mRNA isolation system (Promega, WI) using streptavidin
 Magneticbead particles. Directional cDNA libraries were
 constructed by oligo d(T) priming of poly(A)+ RNA (5mg)
 into EcoRI and XhoI sites of 1 ZapII vector using the Zap
 cDNA synthesis kit (Stratagene, CA). The average size of the
 cDNA inserts in the library was between 1.0 and 1.5kb.
 Clones were mass excised using the Exasist helper phage
 (Stratagene), the phagemids were precipitated with PEG
 8000 and extracted with phenol/chloroform. Phagemid DNA
 was electroporated into DH10B cells."

ORIGIN

Query Match 2.0%; Score 22; DB 13; Length 424;
 Best Local Similarity 100.0%; Pred. No. 55;
 Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 510 TTTAAATAAATTTTAAAA 531
 |||||
 Db 127 TTTAAATAAATTTTAAAA 148
 |||||

Search completed: May 26, 2004, 12:51:08
 Job time : 2588.32 secs